Nonlinear Log-Periodogram Regression for Perturbed Fractional Processes*

Yixiao Sun
Department of Economics
Yale University

Peter C. B. Phillips
Cowles Foundation for Research in Economics
Yale University

First Draft: October 2000
This Version: October 2001

*The authors thank Feng Zhu for his careful reading of the first draft and helpful comments. Sun acknowledges the Cowles Foundation for support under a Carl Anderson fellowship. Phillips thanks the NSF for support under Grant No. SES 0092509. Address correspondence to: Yixiao Sun, Department of Economics, Yale University, P.O. Box 208268, New Haven, CT 06520-8268, USA; Tel.:+1 203 624 6159; Email: yixiao.sun@yale.edu.

ABSTRACT

This paper studies fractional processes that may be perturbed by weakly dependent time series. The model for a perturbed fractional process has a components framework in which there may be components of both long and short memory. All commonly used estimates of the long memory parameter (such as log periodogram (LP) regression) may be used in a components model where the data are affected by weakly dependent perturbations, but these estimates suffer from serious downward bias. To circumvent this problem, the present paper proposes a new procedure that allows for the possible presence of additive perturbations in the data. The new estimator resembles the LP regression estimator but involves an additional (nonlinear) term in the regression that takes account of possible perturbation effects in the data. Under some smoothness assumptions at the origin, the bias of the new estimator is shown to disappear at a faster rate than that of the LP estimator, while its asymptotic variance is inflated only by a multiplicative constant. In consequence, the optimal rate of convergence to zero of the asymptotic MSE of the new estimator is faster than that of the LP estimator. Some simulation results demonstrate the viability and the bias-reducing feature of the new estimator relative to the LP estimator in finite samples.

JEL Classification: C13; C14; C22; C51

Keywords: Asymptotic bias; Asymptotic normality; Bias reduction; Fractional components model; Perturbed fractional process; Rate of convergence

1 Introduction

Fractional processes have been gaining increasing popularity with empirical researchers in economics and finance. In part, this is because fractional processes can capture forms of long run behavior in economic variables that elude other models, a feature that has proved particularly important in modelling inter-trade durations and the volatility of financial asset returns. In part also, fractional processes are attractive to empirical analysts because they allow for varying degrees of persistence, including a continuum of possibilities between weakly dependent and unit root processes.

For a pure fractional process, short run dynamics and long run behavior are driven by the same innovations. This may be considered restrictive in that the innovations that drive long run behavior may arise from quite different sources and therefore differ from those that determine the short run fluctuations of a process. To accommodate this possibility, the model we consider in the present paper allows for perturbations in a fractional process and has a components structure that introduces different sources and types of variation. Such models provide a mechanism for simultaneously capturing the effects of persistent and temporary shocks on the realized observations. They seem particularly realistic in economic and financial applications when there are many different sources of variation in the data and both long run behavior and short run fluctuations need to be modeled.

Specifically, a perturbed fractional process z_t is defined as a fractional process (y_t) that is perturbed by a weakly dependent process (u_t) as follows

$$z_t = y_t + \mu + u_t, t = 1, 2, ..., n, \tag{1}$$

where μ is a constant and

$$y_t = (1-L)^{-d_0} w_t = \sum_{k=0}^{\infty} \frac{\Gamma(d_0+k)}{\Gamma(d_0)\Gamma(k+1)} w_{t-k}, \ 0 < d_0 < 1/2. \tag{2}$$

Here, y_t is a pure fractional process and u_t and w_t are independent Gaussian processes with zero means and continuous spectral densities $f_u(\lambda)$ and $f_w(\lambda)$, respectively. We confine attention to the case where the memory parameter $d_0 \in (0, \frac{1}{2})$ largely for technical reasons that will become apparent later. The case is certainly the most relevant in empirical practice, at least for stationary series, but the restriction is an important one. To maintain generality in the short run components of z_t we do not impose specific functional forms on $f_u(\lambda)$ and $f_w(\lambda)$. Instead, we allow them to belong to a family that is characterized only by regularity conditions near the zero frequency. This formulation corresponds to the conventional semiparametric approach to modelling long range dependence.

By allowing for the presence of two separate stochastic components, the model (1) captures mechanisms in which different factors may come into play in determining long run and short run behaviors. Such mechanisms may be expected to occur in the generation of macroeconomic and financial data for several reasons. For example, time series observations of macroeconomic processes often reflect short run competitive forces as well as long run growth determinants. Additionally, economic and financial time series frequently arise from processes of aggregation and involve errors of measurement, so that the presence of an additive, short memory disturbance is quite realistic. For instance, if the underlying volatility of stock returns follows a fractional process, then realized volatility may follow a perturbed fractional process because the presence of a bid-ask bounce adds a short memory component to realized returns, with consequent effects on volatility.

Some empirical models now in use are actually special cases of perturbed fractional processes. Among these, the long memory stochastic volatility model is growing in popularity for modelling the volatility of financial time series (see Breidt, Crato and De Lima, 1998, and Deo and Hurvich, 1999). This model assumes that $\log r_t^2 = y_t + \mu + u_t$, where r_t is the return, y_t is an underlying fractional process and $u_t = iid(0, \sigma^2)$, thereby coming within the framework of (1). Another example is a rational expectation model in which the ex ante variable follows a fractional process, so that the corresponding ex post variable follows (1) with u_t being a martingale difference sequence. Sun and Phillips (2000) used this framework to model the real rate of interest and inflation as perturbed fractional processes and found that this model helped explain the empirical incompatibility of memory parameter estimates of the components in the ex post Fisher identity. The study by Granger and Marmol (1997) provides a third example, addressing the frequently observed property of financial time series that the autocorrelogram can be low but positive for many lags. Granger and Marmol explained this phenomenon by considering time series that consist of a long memory component combined with a white noise component that has a much larger variance, again coming within the framework of (1).

The main object in the present paper is to develop a suitable estimation procedure for the memory parameter d_0 in (1). As we will show, existing procedures for estimating d_0 typically suffer from serious downward bias in models where there are additive perturbations like (1). The present paper therefore proposes a new procedure that allows for the possible presence of such perturbations in the data.

The spectral density $f_z(\lambda)$ of z_t can be written as $f_z(\lambda) = (2\sin\frac{\lambda}{2})^{-2d_0} f^*(\lambda)$, where $f^*(\lambda) = f_w(\lambda) + (2\sin\frac{\lambda}{2})^{2d_0} f_u(\lambda)$ is a continuous function over $[0, \pi]$. So, $f_z(\lambda)$ satisfies a power law around the origin of the form $f_z(\lambda) \sim G_0 \lambda^{-2d_0}$ as $\lambda \to 0+$, for some positive constant G_0. Therefore, we can estimate d_0 by using the linear log-periodogram (LP) regression introduced by Geweke and Porter-Hudak (1983). Building on the earlier work of Künsch (1986), Robinson (1995a) established the asymptotic normality of the LP estimator. Subsequently, Hurvich, Deo and Brodsky (1998) (hereafter HDB) computed the mean square error of the LP estimator and provided an MSE-optimal rule for bandwidth selection.

The LP estimator has undoubted appeal. It is easy to implement in practice and has been commonly employed in applications. However, when the spectral density of u_t dominates that of w_t in a neighborhood of the origin, the estimator may be biased downward substantially, especially in small samples. One source of the bias is the error of approximating the logarithm of $f^*(\lambda)$ by a constant in a shrinking neighborhood of the origin. This crude approximation also restricts the rate of convergence. The rate of convergence of the LP estimator will be shown to be $n^{-2d_0/(4d_0+1)}$, which is quite slow, especially when d_0 is close to zero.

To alleviate these problems, we take advantage of the structure of our model and propose to estimate the logarithm of $f^*(\lambda)$ locally by $c+\beta\lambda^{2d_0}$. Our new estimator $\widehat{d}$ is defined as the minimizer of the sum of the squared regression errors in a regression of the form

$$\log I_{zj} = \alpha - 2d \log \lambda_j + \beta \lambda_j^{2d} + error, j = 1, 2, ..., m, \tag{3}$$

where

$$I_{zj} = I_z(\lambda_j) = \frac{1}{2\pi n} |\sum_{t=0}^{n-1} z_t \exp(it\lambda_j)|^2, \ \lambda_j = \frac{2\pi j}{n}, \tag{4}$$

and m is a positive integer smaller than the sample size n.

The new estimator can be seen as a way of utilizing parametric information in a nonparametric setting. We approximate the unknown function locally by a nonlinear function instead of a constant. One motivation for the nonlinear LP regression estimator is the local nonlinear least square estimator in the nonparametric literature. Linton and Gozalo (2000) found that the local nonlinear estimator had superior performance compared to the usual kernel estimator when the local nonlinear parameterization is close to the unknown function. Analogously, we expect the nonlinear log periodogram regression estimator to work well in the presence of perturbations, especially when the perturbations are relatively large.

In this paper we investigate the asymptotic and finite sample properties of $\widehat{d}$. We determine its asymptotic bias, variance, asymptotic mean squared error (AMSE), and asymptotic normality, and we calculate the AMSE optimal choice of bandwidth m and its plug-in version. In the presence of the weakly dependent component, we find that the asymptotic bias of $\widehat{d}$ is of order m^{4d_0}/n^{4d_0}, provided that $f_w(\cdot)$ and $f_u(\cdot)$ are boundedly differentiable around the origin, whereas that of the LP estimator $\widehat{d}_{LP}$ has the larger order m^{2d_0}/n^{2d_0}. The asymptotic variances of $\widehat{d}$ and $\widehat{d}_{LP}$ are both of order m^{-1}. In consequence, the optimal rate of convergence to zero of $\widehat{d}$ is of order $n^{-4d_0/(8d_0+1)}$, whereas that of $\widehat{d}_{LP}$ is of the larger order $n^{-2d_0/(4d_0+1)}$. We find that $\widehat{d}$ is asymptotically normal with mean zero, provided that $m^{8d_0+1}/n^{8d_0} \to 0$, whereas $\widehat{d}_{LP}$ is asymptotically normal only under the more stringent condition $m^{4d_0+1}/n^{4d_0} \to 0$.

Some Monte Carlo simulations show that the asymptotic results of the paper mimic the finite sample properties of the new estimator quite well. For the fractional component processes considered in the simulations, the new estimator $\widehat{d}$ has a lower bias, a higher standard deviation, and a lower RMSE compared to the LP estimator $\widehat{d}_{LP}$, as the asymptotic results suggest. The lower bias leads to better coverage probabilities for $\widehat{d}$ over a wide range of m than for $\widehat{d}_{LP}$. On the other hand, the lower standard deviation of $\widehat{d}_{LP}$ leads to shorter confidence intervals than confidence intervals based on $\widehat{d}$.

The paper by Andrews and Guggenberger (1999) is most related to our work. They considered the conventional fractional model (i.e., $var(u_t) = 0$) and proposed to approximate $\log f_w(\lambda)$ by a constant plus a polynomial of even order. Andrews and Sun (2000) investigated the same issue in the context of a local Whittle estimator. Other related papers include Henry and Robinson (1996), Hurvich and Deo (1999) and Henry (1999). These papers consider approximating $\log f^*(\lambda)$ by a more sophisticated function than a constant for the purpose of obtaining a data-driven choice of m. The present paper differs from those papers in that a nonlinear approximation is used in order to achieve bias reduction and to increase the rate of convergence in the estimation of d_0. Also, the nonlinear polynomial function used here depends on the memory parameter d_0 (whereas this is not so in the work just mentioned) and the estimation procedure for d_0 utilizes this information.

The rest of the paper is organized as follows. Section 2 formally defines the new estimator. Section 3 outlines the asymptotics of discrete Fourier transforms and log-periodogram ordinates, which are used extensively in later sections. Section 4 establishes consistency and derives asymptotic normality results for the new estimator. Asymptotic bias, asymptotic MSE, and bandwidth selection are also considered. Section 5 investigates the finite sample performance of the new estimator by simulations. Proofs are collected in the Appendix.

2 Nonlinear Log Periodogram Regression

This section motivates a new estimator that explicitly accounts for the additive perturbations in (1). Throughout, (1) is taken as the data generating process and then

$$f_z(\lambda) = (2\sin\frac{\lambda}{2})^{-2d_0} f^*(\lambda). \tag{5}$$

Taking the logarithms of (5) leads to

$$\log(f_z(\lambda)) = -2d_0\log\lambda + \log f^*(\lambda) - 2d_0\log(2\lambda^{-1}\sin(\frac{\lambda}{2})). \tag{6}$$

Replacing $f_z(\lambda)$ by periodogram ordinates $I_z(\lambda)$ evaluated at the fundamental frequencies $\lambda_j, j = 1, 2, ..., m$ yields

$$\log(I_{zj}) = -c_0 - 2d_0\log\lambda_j + \log f^*(\lambda_j) + U_j + O(\lambda_j^2), \tag{7}$$

where $c_0 = 0.577216...$ is the Euler constant and $U_j = \log[I_z(\lambda_j)/f_z(\lambda_j)] + c_0$.

By virtue of the continuity of $f^*(\lambda)$, we can approximate $\log f^*(\lambda_j)$ by a constant over a shrinking neighborhood of the zero frequency. This motivates log-periodogram regression on the equation

$$\log(I_{zj}) = \text{constant} - 2d\log\lambda_j + error. \tag{8}$$

The LP estimator $\widehat{d}_{LP}$ is then given by the least squares estimator of d in this regression. If $\{U_j\}_{j=1}^m$ behave asymptotically like independent and identically distributed random variables, then the LP estimator is a reasonable choice. In fact, under assumptions to be stated below, we establish that $\sqrt{m}(\widehat{d}_{LP} - d_0) \sim N(b_{LP}, \frac{\pi^2}{24})$ where $b_{LP} = O(m^{2d_0+1}/n^{2d_0})$ and '$\sim$' signifies 'asymptotically distributed.' The 'asymptotic bias' of $\widehat{d}_{LP}$ itself is therefore of order $O(m^{2d_0}/n^{2d_0})$, which can be quite large.

To reduce the bias, we can approximate $\log f^*(\lambda_j)$ by a simple nonlinear function of frequency under the following assumptions:

Assumption 1: *Either (a)* $\sigma_u = var(u_t) = 0$ *for all* t*, so* $f_u(\lambda) \equiv 0$*, for* $\lambda \in [-\pi, \pi]$ *or: (b)* $\sigma_u \neq 0$ *and* $f_u(\lambda)$ *is continuous on* $[-\pi, \pi]$*, bounded above and away from zero with bounded first derivative in a neighborhood of zero.*

Assumption 2: $f_w(\lambda)$ *is continuous on* $[-\pi, \pi]$*, bounded above and away from zero. When* $\sigma_u = 0$*,* $f_w(\lambda)$ *is three times differentiable with bounded third derivative in a neighborhood of zero. When* $\sigma_u \neq 0$*,* $f_w(\lambda)$ *is differentiable with bounded derivative in a neighborhood of zero.*

Assumptions 1(b) and 2 are local smoothness conditions and hold for many models in current use, including ARMA models. They allow us to develop a Taylor expansion of $\log f^*(\lambda)$ about $\lambda = 0$ with an error of the order of the first omitted term. Specifically, when $\sigma_u = 0$,

$$\log f^*(\lambda_j) = \log f_w(0) + O(\lambda_j^2). \tag{9}$$

When $\sigma_u \neq 0$,

$$\begin{aligned} & \log f^*(\lambda_j) \\ = \; & \log f_w(\lambda_j) + \log[1 + (2\sin\frac{\lambda_j}{2})^{2d_0}\frac{f_u(\lambda_j)}{f_w(\lambda_j)}] \\ = \; & \log f_w(\lambda_j) + \log\left\{1 + \lambda_j^{2d_0}(1 + O(\lambda_j^2))\left(\frac{f_u(0)}{f_w(0)} + O(\lambda_j^2)\right)\right\} \\ = \; & \log f_w(0) + \frac{f_u(0)}{f_w(0)}\lambda_j^{2d_0} + O(\lambda_j^{4d_0}). \end{aligned} \tag{10}$$

So, in either case

$$\log f^*(\lambda_j) = \log f_w(0) + \frac{f_u(0)}{f_w(0)}\lambda_j^{2d_0} + O(\lambda_j^r) \tag{11}$$

where $O(\cdot)$ holds uniformly over $j = 1, 2, ..., m$ and $r = 4d_0\{\sigma_u \neq 0\} + 2\{\sigma_u = 0\}$.

Combining (7) with (11) produces the nonlinear LP regression model:

$$\log(I_{zj}) = -2d_0 \log\lambda_j + \alpha_0 + \lambda_j^{2d_0}\beta_0 + U_j + \varepsilon_j, \tag{12}$$

where

$$\begin{aligned} \alpha_0 &= \log f_w(0) - c_0, \beta_0 = f_u(0)/f_w(0), \text{ and} \\ \varepsilon_j &= \log f^*(\lambda_j) - \log f_w(0) - \beta_0\lambda_j^{2d_0} - 2d_0[\log(2\sin\frac{\lambda_j}{2}) - \log\lambda_j]. \end{aligned} \tag{13}$$

The new estimator is then defined as the minimizer of the sum of squared regression errors in this model, i.e.

$$(\hat{\alpha}, \widehat{d}, \widehat{\beta}) = \arg\min_{\alpha,d,\beta} SSE(\alpha, d, \beta), \tag{14}$$

where

$$SSE(\alpha, d, \beta) = \sum_{j=1}^{m}[\log(I_{zj}) - \alpha + 2d \log \lambda_j - \lambda_j^{2d}\beta]^2. \tag{15}$$

Concentrating (15) with respect to α, we obtain

$$(\widehat{d}, \widehat{\beta}) = \arg \min_{d \in D, \beta \in B} Q(d, \beta), \tag{16}$$

with

$$\begin{aligned} Q(d, \beta) &= \sum_{j=1}^{m}\{(\log I_{zj} - \frac{1}{m}\sum_{k=1}^{m} \log I_{zk}) \\ &\quad +2d(\log \lambda_j - \frac{1}{m}\sum_{k=1}^{m} \log \lambda_k) - \beta(\lambda_j^{2d} - \frac{1}{m}\sum_{k=1}^{m} \lambda_k^{2d})\}^2. \end{aligned} \tag{17}$$

where B is a compact and convex set, $D = [d_1, d_2]$ is a closed interval of admissible values for d_0 with $0 < d_1 < d_2 < 1/2$. Here d_1 and d_2 can be chosen arbitrarily close to 0 and 1/2, respectively. We write $\theta = (d, \beta)$, $\Theta = D \otimes B$ for convenience and assume the true value of θ lies in the interior of the admissible set.

3 Log-periodogram Asymptotics and Useful Lemmas

To establish the asymptotic properties of the new estimator, we need to characterize the asymptotic behavior of the log-periodogram ordinates $U_j = \log[I_z(\lambda_j)/f_z(\lambda_j)] + c_0$. Define

$$A_{zj} = \frac{1}{\sqrt{2\pi n}} \sum_{t=0}^{n-1} z_t \cos \lambda_j t \text{ and } B_{zj} = \frac{1}{\sqrt{2\pi n}} \sum_{t=0}^{n-1} z_t \sin \lambda_j t, \tag{18}$$

then

$$U_j = \ln\left(\frac{A_{zj}^2}{f_{zj}} + \frac{B_{zj}^2}{f_{zj}}\right) + c_0, j = 1, ..., m. \tag{19}$$

In view of the Gaussianity of A_{zj} and B_{zj}, we can evaluate the means, variances, and covariances of U_j, if the asymptotic behavior of the vector $\left(A_{zj}/f_{zj}^{1/2}, B_{zj}/f_{zj}^{1/2}, A_{zk}/f_{zk}^{1/2}, B_{zk}/f_{zk}^{1/2}\right)$ is known. The properties of this vector depend in turn on those of the discrete Fourier transforms of z_t, defined as $w(\lambda) = (2\pi n)^{-1/2} \sum_1^n z_t e^{it\lambda}$.

The asymptotic behavior of $w(\lambda)$ is given in the following lemma which is a variant of results given earlier by several other authors (Robinson, 1995a, HDB, 1998, Andrews and Guggenberger, 1999).

Lemma 1 *Let Assumptions 1 and 2 hold. Then uniformly over j and k, $1 \leq k < j \leq m$, $m/n \to 0$,*

(a) $E\left[w\left(\lambda_j\right)\overline{w}\left(\lambda_j\right)/f_z\left(\lambda_j\right)\right]=1+O\left(j^{-1}\log j\right),$

(b) $E\left[w\left(\lambda_j\right)w\left(\lambda_j\right)/f_z\left(\lambda_j\right)\right]=O\left(j^{-1}\log j\right),$

(c) $E\left[w\left(\lambda_j\right)\overline{w}\left(\lambda_k\right)/\left(f_z\left(\lambda_j\right)f_z\left(\lambda_k\right)\right)^{1/2}\right]=O\left(k^{-1}\log j\right),$

(d) $E\left[w\left(\lambda_j\right)w\left(\lambda_k\right)/\left(f_z\left(\lambda_j\right)f_z\left(\lambda_k\right)\right)^{1/2}\right]=O\left(k^{-1}\log j\right).$

It follows directly from Lemma 1 that for $1\leq k<j\leq m$,

$$\begin{aligned} EA_{zj}^2/f_{zj} &= \frac{1}{2}+O(\frac{\log j}{j}),\ EB_{zj}^2/f_{zj}=\frac{1}{2}+O(\frac{\log j}{j}),\\ EA_{zj}B_{zj}/f_{zj} &= O(\frac{\log j}{j}),\ EA_{zj}B_{zk}/(f_{zj}f_{zk})^{1/2}=O(\frac{\log j}{k}). \end{aligned} \tag{20}$$

Using these results and following the same line of derivation as in HDB (1998), we can prove Lemma 2 below. Since the four parts of this lemma are proved in a similar way to Lemmas 3, 5, 6 and 7 in HDB, the proofs are omitted here.

Lemma 2 *Let Assumptions 1 and 2 hold. Then*

(a) $Cov\left(U_j,U_k\right)=O\left(\log^2 j/k^2\right),$ *uniformly for* $\log^2 m\leq k<j\leq m,$

(b) $\lim_n \sup_{1\leq j\leq m} EU_j^2<\infty,$

(c) $E\left(U_j\right)=O\left(\log j/j\right),$ *uniformly for* $\log^2 m\leq j\leq m,$

(d) $Var\left(U_j\right)=\pi^2/6+O\left(\log j/j\right),$ *uniformly for* $\log^2 m\leq j\leq m.$

With the asymptotic behavior of U_j in hand, we can proceed to show that the normalized sums $\frac{1}{m}\sum_{j=1}^m c_jU_j$ are uniformly negligible under certain conditions on the coefficients c_j. Quantities of this form appear in the normalized Hessian matrix below.

Lemma 3 *Let* $\{c_j(d,\beta)\}_{j=1}^m$ *be a sequence of functions such that, for some* $p\geq 0,$

$$\sup_{(d,\beta)\in\Theta}|c_j|=O(\log^p m)\ \textit{uniformly for}\ 1\leq j\leq m, \tag{21}$$

and for some $q\geq 0,$

$$\sup_{(d,\beta)\in\Theta}|c_j-c_{j-1}|=O(j^{-1}\log^q m)\ \textit{uniformly for}\ 1\leq j\leq m. \tag{22}$$

Then

$$\sup_{(d,\beta)\in\Theta}\left|\frac{1}{m}\sum_{j=1}^m c_jU_j\right|=O_p(\frac{\log^{\max(p,q)}m}{\sqrt{m}}). \tag{23}$$

We can impose additional conditions to get a tighter bound. For example, if we also require that $\sup_{(d,\beta)\in\Theta}|c_m| = O(1)$, then $\sup_{(d,\beta)\in\Theta}\left|\frac{1}{m}\sum_{j=1}^{m} c_j U_j\right| = O_p(\frac{\log^q m}{\sqrt{m}})$, as is readily seen from the proof of the lemma. Further, the lemma remains valid if we remove the 'sup' operator from both the conditions and the conclusion.

The following lemma assists in establishing the asymptotic normality of the nonlinear log-periodogram regression estimator.

Lemma 4 *Let $a_{kn} = a_k$ be a triangular array for which*

$$\max_k |a_k| = o(m), \qquad \sum_{k=1+m^{0.5+\delta}}^{m} a_k^2 \sim \rho m, \qquad \sum_{k=1+m^{0.5+\delta}}^{m} |a_k|^p = O(m), \tag{24}$$

for all $p \geq 1$, and $0 < \delta < 0.5$. Then,

$$\frac{1}{\sqrt{m}} \sum_{k=1+m^{0.5+\delta}}^{m} a_k U_k \xrightarrow{d} N\left(0, \frac{\pi^2}{6}\rho\right). \tag{25}$$

The proof of this lemma is based on the method of moments and involves a careful exploration of the dependence structure of the discrete Fourier transforms. Robinson's argument (1995a, pp. 1067-70) forms the basis of this development and can be used here with some minor modifications to account for differences in the models. Details are omitted here and are available upon request.

4 Consistency, Asymptotic Normality and Bandwidth Choice

We first establish asymptotic properties for the LP estimator in the context of the components model (1). The following theorem gives the limit theory and provides a benchmark for later comparisons.

Theorem 1 *Let Assumptions 1 and 2 hold. Let $m = m(n) \to \infty$ and*

$$\frac{m^{2r'+1}}{n^{2r'}} \to K_\sigma\{\sigma_u \neq 0\} + K_0\{\sigma_u = 0\} \tag{26}$$

as $n \to \infty$, where $r' = 2d_0\{\sigma_u \neq 0\} + 2\{\sigma_u = 0\}$ and $K_\sigma, K_0 > 0$ are constants. Then

$$\sqrt{m}(\widehat{d}_{LP} - d_0) \Rightarrow N(b_{LP}, \frac{\pi^2}{24}), \tag{27}$$

where

$$b_{LP} = -(2\pi)^{2d_0} \frac{f_u(0)}{f_w(0)} \frac{d_0}{(2d_0+1)^2} K_\sigma\{\sigma_u \neq 0\} - \frac{2\pi^2}{9}\left(\frac{f_w''(0)}{f_w(0)} + \frac{d_0}{6}\right) K_0\{\sigma_u = 0\}. \tag{28}$$

When $\sigma_u \neq 0$, the ratio $m^{2r'+1}/n^{2r'} = m^{4d_0+1}/n^{4d_0} \to K_\sigma$ in (26). This delivers an upper bound of order $O(n^{4d_0/(1+4d_0)})$ on the rate at which m can increase with n and allows for larger choices of m for larger values of d_0. Intuitively, as d_0 increases, the contamination from perturbations at frequencies away from the origin becomes relatively smaller and we can expect to be able to employ a wider bandwidth in the regression. To eliminate the asymptotic bias b_{LP} in (27) altogether, we use a narrower band and set $m = o(n^{4d_0/(1+4d_0)})$ in place of (26). Deo and Hurvich (1999) established a similar result under the assumption that u_t is iid, but not necessarily Gaussian. Their assumption that $m^{4d_0+1}\log^2 m/n^{4d_0} = o(1)$ is slightly stronger than the assumption made here.

When $\sigma_u \neq 0$, the limit distribution (27) involves the bias

$$b_{LP} = -(2\pi)^{2d_0}\frac{f_u(0)}{f_w(0)}\frac{d_0}{(2d_0+1)^2}K_\sigma < 0, \tag{29}$$

which is always negative, as one would expect, because of the effect of the short memory perturbations. Correspondingly, the dominating bias term of $\widehat{d}_{LP}$ has the form

$$b_{n,LP} = -(2\pi)^{2d_0}\frac{f_u(0)}{f_w(0)}\frac{d_0}{(2d_0+1)^2}\frac{m^{2d_0}}{n^{2d_0}} < 0. \tag{30}$$

The magnitude of the bias obviously depends on the quantity $f_w(0)/f_u(0)$, which is the ratio of the long run variance of the short memory input of y_t to that of the perturbation component u_t. The ratio can be interpreted as a long run signal-noise ratio (SNR), measuring the strength in the long run of the signal from the y_t inputs relative to the long run signal in the perturbations. The stronger the long run signal in the perturbations, the greater the downward bias and the more difficult it becomes to estimate the memory parameter accurately. One might expect these effects to be exaggerated in small samples where the capacity of the data to discriminate between long run and short run effects is reduced.

When $\sigma_u = 0$, the theorem contains essentially the same results proved in HDB. In this case, the dominating term in the bias of $\widehat{d}_{LP}$ is given by

$$b_{n,LP} = -\frac{2\pi^2}{9}\left(\frac{f_w''(0)}{f_w(0)} + \frac{d_0}{6}\right)\frac{m^2}{n^2}. \tag{31}$$

HDB showed that the dominating bias of $\widehat{d}_{LP}$ in the case of pure fractional process regression is given by the expression

$$-\frac{2\pi^2}{9}\left(\frac{f_w''(0)}{f_w(0)}\right)(\frac{m}{n})^2. \tag{32}$$

The presence of the additional factor $d_0/6$ in the second term of our expression (31) arises from the use of a slightly different regressor in the LP regression. In particular, we employ $-2\log\lambda_j$ as one of the regressors in (3), while HDB use $-2\log(2\sin\lambda_j/2)$. These regressors are normally considered to be asymptotically equivalent. However, while the use of $-2\log\lambda_j$ rather than $-2\log(2\sin\lambda_j/2)$ has no effect on the asymptotic variance, it does affect the asymptotic bias.

We now investigate the asymptotic properties of the nonlinear log-periodogram regression estimator.

Theorem 2 *Let Assumptions 1 and 2 hold.*

(a) *If* $\frac{1}{m} + \frac{m}{n} \to 0$ *as* $m, n \to \infty$, *then* $\widehat{d} - d_0 = o_p(1)$.

(b) *If for some arbitrary small* $\Delta > 0, \frac{m}{n} + \frac{n^{4d_0(1+\Delta)}}{m^{4d_0(1+\Delta)+1}} \to 0$, *as* $m, n \to \infty$, *then* $\widehat{d} - d_0 = O_p\left((\frac{m}{n})^{2d_0}\right)$ *and* $\widehat{\beta} - \beta_0 = o_p(1)$.

Theorem 2 shows that $\widehat{d}$ is consistent under mild conditions. All that is needed is that m approaches infinity slower than the sample size n. As shown by HDB, trimming out low frequencies is not necessary. This point is particularly important in the present case. In seeking to reduce contaminations from the perturbations, the lowest frequency ordinates are the most valuable in detecting the long memory effects.

It is not straightforward to establish the consistency of $\widehat{\beta}$, because, as $n \to \infty$, the objective function becomes flat as a function of β. The way we proceed is, in fact, to show first that $\widehat{d}$ converges to d_0 at some slower rate, more precisely, $\widehat{d} - d_0 = O_p((\frac{m}{n})^{2d_0})$. We prove this rate of convergence stepwise. We start by showing that $\widehat{d} - d_0 = o_p((\frac{m}{n})^{d_1/2})$ for $0 < d_1 < d_0$, using the fact that $\beta\lambda_j^{2d} = O(\frac{m}{n})^{2d_1}$ uniformly in $(d, \beta) \in \Theta$. We can then deduce that $\widehat{d} - d_0 = o_p((\frac{m}{n})^{d_0(1+\Delta)})$. With this faster rate of convergence, we have better control over some quantities and can obtain an even faster rate of convergence for $\widehat{d}$. Repeating this procedure leads to $\widehat{d} - d_0 = O_p((\frac{m}{n})^{2d_0})$, as desired. The idea of the proof may be applicable to other nonlinear estimation problems when the involved variables are integrated of different orders or have different stochastic orders.

We prove the rate of convergence of $\widehat{d}$ without using the consistency of $\widehat{\beta}$. This is unusual because in most nonlinear estimation problems it is common to prove the consistency of all parameters first in order to establish rates of convergence. The approach is successful in the present case because when d is close to d_0, the regressor λ_j^{2d} evaporates as $n \to \infty$ and approaches zero approximately at the rate of $(m/n)^{2d_0}$.

We proceed to show that under somewhat stronger conditions and if suitably normalized, $(\widehat{d} - d_0, \widehat{\beta} - \beta_0)$ is asymptotically normal. The new assumption is as follows:

Assumption 3: $n^{4d_0(1+\Delta)}/m^{4d_0(1+\Delta)+1} \to 0$ *for some arbitrary small* $\Delta > 0$ *and* $m^{2r+1}/n^{2r} = O(1)$ *as* $m, n \to \infty$, *where* $r = 4d_0\{\sigma_u \neq 0\} + 2\{\sigma_u = 0\}$.

The two conditions in Assumption 3 are always compatible because $r \geq 4d_0$ and Δ is arbitrarily small. The lower bound on the growth rate of m ensures the consistency of $\widehat{d}$ and $\widehat{\beta}$, which validates the use of the first order conditions. The upper bound on the growth rate of m guarantees that the normalized gradient of $Q(d, \beta)$ is $O_p(1)$, which is required for the asymptotic normality of $(\widehat{d}, \widehat{\beta})$.

When $\sigma_u = 0$, the upper bound becomes $m^5/n^4 = O(1)$, which is the same as the upper bound for asymptotic normality of the LP estimator for a pure fractional process.

When $\sigma_u \neq 0$, the upper bound becomes $m^{8d_0+1}/n^{8d_0} = O(1)$, which is less stringent than the upper bound given in Theorem 1. It therefore allows us to take m larger than in conventional LP regression applied to the fractional components model. In consequence, by an appropriate choice of m, we have asymptotic normality for $\widehat{d}$ with a faster rate of convergence than is possible in LP regression. However, for any $0 < d_0 < 1/2$, the upper bound is more stringent than $m = O(n^{4/5})$, the upper bound for asymptotic normality of LP regression in a pure fractional process model. Hence, the existence of the weakly dependent perturbations in (1) requires the use of a narrower bandwidth than LP regression for a pure fractional process. Interestingly, as d_0 approaches 1/2, the upper bound becomes arbitrarily close to $m = O(n^{4/5})$.

We now proceed to establish asymptotic normality. The first order conditions for (16) are:

$$S_n(\widehat{d}, \widehat{\beta}) = 0, \tag{33}$$

where

$$S_n(d,\beta) = -\sum_{j=1}^{m} \begin{pmatrix} x_{1j}(d,\beta) - \bar{x}_1(d,\beta) \\ x_{2j}(d,\beta) - \bar{x}_2(d,\beta) \end{pmatrix} e_j(d,\beta), \tag{34}$$

$$\begin{aligned} x_{1j}(d,\beta) &= -2\log\lambda_j(1-\beta\lambda_j^{2d}), \bar{x}_1(d,\beta) = \frac{1}{m}\sum_{k=1}^{m} x_{1k}, \\ x_{2j}(d,\beta) &= \lambda_j^{2d}, \bar{x}_2(d,\beta) = \frac{1}{m}\sum_{k=1}^{m} x_{2k}, \text{ and} \end{aligned} \tag{35}$$

$$e_j(d,\beta) = \log I_{zj} - \frac{1}{m}\sum_{k=1}^{m}\log I_{zk} + 2d(\log\lambda_j - \frac{1}{m}\sum_{k=1}^{m}\log\lambda_k) - \left(\beta\lambda_j^{2d} - \frac{1}{m}\sum_{k=1}^{m}\beta\lambda_k^{2d}\right). \tag{36}$$

Expanding $S_n(\widehat{d}, \widehat{\beta})$ about $S_n(d_0, \beta_0)$, we have

$$0 = S_n(d_0,\beta_0) + H_n(d_0,\beta_0)(\widehat{d}-d_0, \widehat{\beta}-\beta_0)' + [H_n^* - H_n(d_0,\beta_0)](\widehat{d}-d_0, \widehat{\beta}-\beta_0)', \tag{37}$$

where H_n is the Hessian matrix, H_n^* is the Hessian evaluated at (d^*, β^*), the mean values between (d_0, β_0) and $(\widehat{d}, \widehat{\beta})$. The elements of the Hessian matrix are:

$$\begin{aligned} H_{n,11}(d,\beta) &= \sum_{j=1}^{m}(x_{1j} - \bar{x}_1)^2 - \beta\sum_{j=1}^{m} e_j \left(\log\lambda_j^2\right)^2 \lambda_j^{2d}, \\ H_{n,12}(d,\beta) &= \sum_{j=1}^{m}(x_{1j} - \bar{x}_1)(x_{2j} - \bar{x}_2) - \sum_{j=1}^{m} e_j \left(\log\lambda_j^2\right) \lambda_j^{2d}, \\ H_{n,22}(d,\beta) &= \sum_{j=1}^{m}(x_{2j} - \bar{x}_2)^2. \end{aligned} \tag{38}$$

Define the diagonal matrix $D_n = diag(\sqrt{m}, \lambda_m^{2d_0}\sqrt{m})$. We show in the following lemma that the normalized Hessian $D_n^{-1}H_n(d_0, \beta_0)D_n^{-1}$ converges in probability to a 2×2 matrix defined by

$$\Omega = \begin{pmatrix} 4 & \frac{-4d_0}{(2d_0+1)^2} \\ \frac{-4d_0}{(2d_0+1)^2} & \frac{4d_0^2}{(4d_0+1)(2d_0+1)^2} \end{pmatrix}, \tag{39}$$

and the 'asymptotic bias' of the normalized score $D_n^{-1}S_n(d_0, \beta_0)$ is $-b_n$, where

$$b_n = \{\sigma_u \neq 0\} m^{1/2} \lambda_m^{4d_0} b_{1n} + \{\sigma_u = 0\} m^{1/2} \lambda_m^2 b_{2n}, \tag{40}$$

and

$$b_{1n} = \frac{f_w^2(0)}{2f_u^2(0)} \begin{pmatrix} \frac{8d_0}{(4d_0+1)^2} \\ -\frac{8d_0^2}{(2d_0+1)(4d_0+1)(6d_0+1)} \end{pmatrix},$$

$$b_{2n} = \left(\frac{f_w''(0)}{f_w(0)} + \frac{d_0}{6}\right) \begin{pmatrix} -\frac{2}{9} \\ \frac{2d_0}{3(2d_0+3)(2d_0+1)} \end{pmatrix}. \tag{41}$$

Before stating the lemma, we need the following notation. Let $J_n(d, \beta)$ be a 2×2 matrix whose (i, j)-th element is

$$J_{n,ij} = \sum_{k=1}^{m} \left(x_{ik}(d, \beta) - \bar{x}_i(d, \beta)\right) \left(x_{jk}(d, \beta) - \bar{x}_j(d, \beta)\right), \tag{42}$$

and let Θ_n be a set defined by

$$\Theta_n = \{(d, \beta) : |\lambda_m^{-d_0}(d - d_0)| < \varepsilon \text{ and } |\beta - \beta_0| < \varepsilon\}. \tag{43}$$

Lemma 5 *Let Assumptions 1-3 hold. We have*

(a) $\sup_{(d,\beta)\in\Theta_n} ||D_n^{-1}(H_n(d, \beta) - J_n(d, \beta))D_n^{-1}|| = o_p(1)$,

(b) $\sup_{(d,\beta)\in\Theta_n} ||D_n^{-1}[J_n(d, \beta) - J_n(d_0, \beta_0)]D_n^{-1}|| = o_p(1)$,

(c) $D_n^{-1}J_n(d_0, \beta_0)D_n^{-1} \to \Omega$,

(d) $D_n^{-1}S_n(d_0, \beta_0) + b_n \Rightarrow N(0, \frac{\pi^2}{6}\Omega)$.

Theorem 3 *Let Assumptions 1, 2 and 3 hold, then*

$$D_n \begin{pmatrix} \widehat{d} - d_0 \\ \widehat{\beta} - \beta_0 \end{pmatrix} - \Omega^{-1} b_n \Rightarrow N(0, \frac{\pi^2}{6}\Omega^{-1}) \tag{44}$$

where

$$\Omega^{-1} = \begin{bmatrix} \frac{1}{16d_0^2}(2d_0+1)^2 & \frac{1}{16d_0^3}(2d_0+1)^2(4d_0+1) \\ \frac{1}{16d_0^3}(2d_0+1)^2(4d_0+1) & \frac{1}{16d_0^4}(4d_0+1)(2d_0+1)^4 \end{bmatrix}.$$

Remark 1 From the above theorem, we deduce immediately that the asymptotic variance of $\sqrt{m}(\widehat{d}-d_0)$ is $\frac{\pi^2}{24}C_d$, where $C_d = 1+\frac{4d_0+1}{4d_0^2} > 1$. Approximating $\log f^*(\cdot)$ locally by a nonlinear function instead of a constant therefore inflates the usual asymptotic variance of the LP regression estimator in a pure fractional model by the factor C_d. This is to be expected, as adding more variables in regression usually inflates variances.

Remark 2 The 'asymptotic bias' of $(\widehat{d},\widehat{\beta})'$ is equal to $D_n^{-1}\Omega^{-1}b_n$. Some algebraic manipulations show that when $\sigma_u = 0$,

$$D_n^{-1}\Omega^{-1}b_n = -\frac{2\pi^2}{9}\left(\frac{f_w''(0)}{f_w(0)}+\frac{d_0}{6}\right)(\frac{m}{n})^2\begin{pmatrix}\frac{(d_0-1)(2d_0+1)}{d_0(2d_0+3)}\\ \frac{(2d_0+1)^2(4d_0+1)}{d_0^2(2d_0+3)}\end{pmatrix}, \tag{45}$$

and when $\sigma_u \neq 0$,

$$D_n^{-1}\Omega^{-1}b_n = -\frac{(2\pi)^{4d_0}f_w^2(0)}{f_u^2(0)}(\frac{m}{n})^{4d_0}\begin{pmatrix}\frac{d_0(2d_0+1)}{(4d_0+1)^2(6d_0+1)}\\ \frac{2(2d_0+1)^2}{(4d_0+1)(6d_0+1)}\end{pmatrix}. \tag{46}$$

Remark 3 When $\sigma_u \neq 0$, according to (46) the asymptotic bias of $\widehat{d}$ is of order m^{4d_0}/n^{4d_0}. In contrast, the asymptotic bias of the LP estimator is of order m^{2d_0}/n^{2d_0}, as shown above in (30). The asymptotic bias of the new estimator is therefore smaller than that of the LP estimator by order m^{2d_0}/n^{2d_0}. When $\sigma_u = 0$, the asymptotic bias of $\widehat{d}$ is of the same order as that of $\widehat{d}_{LP}$, as seen from (31) and (45). The relative magnitude depends on the value of d_0 and the curvature of $f_w(\lambda)$ at $\lambda = 0$.

Remark 4 Note that $\widehat{\beta}$ converges more slowly by a rate of $(\frac{m}{n})^{-2d_0}$ than $\widehat{d}$. Heuristically, the excitation levels of the two regressors ($\log\lambda_j$ and $\lambda_j^{2d_0}$) and thus their information content are different. More specifically, we have $\sum_{j=1}^m(\log\lambda_j - \sum_{k=1}^m \log\lambda_k/m)^2 = O(m)$ whereas $\sum_{j=1}^m(\lambda_j^{2d_0}-\sum_{k=1}^m\lambda_k^{2d_0}/m)^2 = O(m\lambda_m^{2d_0})$.

Next, we consider issues of bandwidth choice in the case where $\sigma_u \neq 0$. Following the above remarks, the asymptotic mean-squared error (AMSE) of $\widehat{d}$ is

$$AMSE(\widehat{d}) = K^2(\frac{m}{n})^{8d_0} + \frac{\pi^2}{24m}C_d, \tag{47}$$

where

$$K = (2\pi)^{4d_0}\beta_0^2\frac{d_0\,(2d_0+1)}{(4d_0+1)^2\,(6d_0+1)}. \tag{48}$$

Straightforward calculations yield the value of m that minimizes $AMSE(\widehat{d})$, viz.

$$m^{opt} = [(\frac{\pi^2C_d}{192d_0K^2})^{1/(8d_0+1)}n^{8d_0/(8d_0+1)}], \tag{49}$$

where $[\cdot]$ denotes the integer part.

In contrast, the AMSE of $\widehat{d}_{LP}$ is

$$AMSE(\widehat{d}_{LP}) = K_{LP}^2 (\frac{m}{n})^{4d_0} + \frac{\pi^2}{24m}, \tag{50}$$

where

$$K_{LP} = (2\pi)^{2d_0} \beta_0 \frac{d_0}{(2d_0+1)^2}. \tag{51}$$

So the AMSE-optimal bandwidth for $\widehat{d}_{LP}$ is

$$m_{LP}^{opt} = [(\frac{\pi^2}{96 d_0 K_{LP}^2})^{1/(4d_0+1)} n^{4d_0/(4d_0+1)}]. \tag{52}$$

When $m = m^{opt}$, the AMSE of $\widehat{d}$ converges to zero at the rate of $n^{-8d_0/(8d_0+1)}$. In contrast, when $m = m_{LP}^{opt}$, the AMSE of $\widehat{d}_{LP}$ converges to zero only at the rate of $n^{-4d_0/(4d_0+1)}$. Thus, the optimal AMSE of $\widehat{d}$ converges to zero faster than that of $\widehat{d}_{LP}$.

Of course, the optimal bandwidth (49) depends on the unknown quantities d_0 and β_0. Consistent estimates are readily available under Assumptions 1, 2 and 3. A data dependent choice of m for the computation of $\widehat{d}$ can be obtained by plugging initial estimates of β_0 and d_0 into (49).

5 Simulations

5.1 Experimental Design

This section investigates the finite sample performance of the new estimator in comparison with conventional LP regression. The chosen data generating process is

$$z_t = (1-L)^{-d_0} w_t + u_t, \tag{53}$$

where $\{w_t : t = 1, 2, ..., n\}$ are iid $N(0,1)$, $\{u_t : t = 1, 2, ..., n\}$ are iid $N(0, \sigma_u^2)$ and $\{w_t\}$ are independent of $\{u_t\}$.

We consider the following constellation of parameter combinations

$$\begin{aligned} d_0 &= 0.25, 0.45, 0.65, 0.85, \text{ and} \\ \sigma_u^2 &= 0, 4, 8, 16. \end{aligned} \tag{54}$$

In view of the fact that the LP estimator is consistent for both stationary fractional processes ($d_0 < 0.5$) and nonstationary fractional processes ($0.5 \leq d_0 < 1$) (see Kim and Phillips, 1999), we expect the new estimator to work well for nonstationary fractional component processes for this range of values of d_0 as well as for stationary fractional component processes over ($0 < d_0 < 0.5$). Hence it is of interest to include some values of d_0 that fall in the nonstationary zone.

The value of σ_u^2 determines the strength of the noise from the perturbations. The long run SNR increases as σ_u^2 decreases. When $\sigma_u^2 = 0$, z_t is a pure fractional process with an infinite long run SNR. The inverse of the long run SNR, viz. $f_u(0)/f_w(0)$, takes the values $0, 4, 8, 16$. These are close to the values in Deo and Hurvich (1999). In their simulation study, the ratio $f_u(0)/f_w(0)$ takes the values 6.17 and 13.37.

We consider sample sizes $n = 512$ and 2048. Because n has the composite form 2^k (k integer) for these choices, zero-padding is not a concern when we use the fast Fourier transform to compute the periodogram. For each sample size and parameter combination, 2000 replications are performed from which we calculate the biases, standard deviations and root mean square errors of $\widehat{d}$ and $\widehat{d}_{LP}$, for different selections of the bandwidth m. Then, for each parameter combination, we graph each of these quantities as functions of m ($m = 4, 5, ..., n/2$). The results are shown in panels (a)-(c) of Figs. 1–5.

In addition, we compute the coverage probabilities, as functions of m, of the nominal 90% confidence intervals that are obtained using the asymptotic normality results of Theorems 1 and 3. When constructing these confidence intervals, we estimate the variances of $\widehat{d}$ and $\widehat{d}_{LP}$ using finite sample expressions rather than the limit expressions, because the former yield better finite sample performance for all parameter combinations and for both estimators. The variance of $\widehat{d}$ is estimated by the (1,1) element of the inverse of the Hessian matrix, which is

$$\frac{\pi^2}{6} H_{22,n}(\widehat{d}, \widehat{\beta}) \left(H_{11}(\widehat{d}, \widehat{\beta}) H_{22}(\widehat{d}, \widehat{\beta}) - H_{12}^2(\widehat{d}, \widehat{\beta}) \right)^{-1}, \tag{55}$$

whereas the variance of $\widehat{d}_{LP}$ is estimated by

$$\frac{\pi^2}{24} \left(\sum_{j=1}^{m} \log \lambda_j - \frac{1}{m} \sum_{k=1}^{m} \log \lambda_k \right)^{-2}. \tag{56}$$

We calculate the average lengths of the confidence intervals as functions of m. The coverage probabilities and the average lengths are graphed against m in panels (d) and (e) of Figs. 1–5.

5.2 Results

We report results only for the cases $d_0 = 0.45$ and $d_0 = 0.85$, since these are representative of the results found in the other two cases, $d_0 = 0.25$ and 0.65, respectively. Also, for each value of d_0, we discuss only the cases $\sigma_u^2 = 0$ and $\sigma_u^2 = 8$, as the results for the other values of σ_u^2 were qualitatively similar.

We first discuss the results when $d_0 = 0.45$ and $\sigma_u^2 = 0$. In this case, z_t is a pure fractional process. Fig. 1(a) shows that the bias of $\widehat{d}$ is positive and larger than that of $\widehat{d}_{LP}$. The positive bias of $\widehat{d}$ conforms to our asymptotic results. From Remark 2, the asymptotic bias of $\widehat{d}$ is $-\frac{\pi^2}{27}(\frac{m}{n})^2(d_0 - 1)(2d_0 + 1)(2d_0 + 3)^{-1}$, which is always positive for $d_0 < 1$. Fig. 1(b) shows that the variance of $\widehat{d}$ is larger than that of $\widehat{d}_{LP}$, as predicted by Theorem 3. Comparing RMSE's in Fig. 1(c), we see that the RMSE of $\widehat{d}$ is larger than that of $\widehat{d}_{LP}$. The inferior performance of $\widehat{d}$ in this case is not surprising since the LP estimator is designed for pure fractional processes, whereas our estimator $\widehat{d}$ allows for additional noise in the system and is designed for perturbed fractional processes. However, it is encouraging that the LP estimator outperforms the new estimator only by a small margin. Apparently, the cost of including the additional regressor, even when it is not needed, is small.

Next, we discuss the results when $d_0 = 0.45$ and $\sigma_u^2 = 8$. Fig. 2(a) shows that the LP estimator $\widehat{d}_{LP}$ has a large downward bias in this case, whereas the new estimator $\widehat{d}$ has a much smaller bias. Apparently, the bias-reducing feature of $\widehat{d}$ established in the asymptotic theory is manifest in finite samples. Fig. 2(b) shows that the standard error of $\widehat{d}_{LP}$ is less than that of $\widehat{d}$ for all values of m, again consistent with the asymptotic results. For each estimator, the standard error declines at the approximate rate $1/\sqrt{m}$ as m increases, because m is the effective sample size in the estimation of d_0. Fig. 2(c) shows that the RMSE of $\widehat{d}$ is smaller than that of $\widehat{d}_{LP}$ over a wide range of m values. Fig. 2(d) shows that the coverage probability of $\widehat{d}$ is fairly close to the nominal value of 0.9, provided that m is not taken too large. In contrast, $\widehat{d}_{LP}$ has a true coverage probability close to 0.9 only for very small values of m. This is due to the large bias of $\widehat{d}_{LP}$. However, the larger standard error of $\widehat{d}$ leads to longer confidence intervals on average, and this is apparent in Fig. 2(e).

We now turn to the results when $d_0 = 0.85$ and $\sigma_u^2 = 0$. Figure 3 shows that both $\widehat{d}_{LP}$ and $\widehat{d}$ work reasonably well for nonstationary fractional processes $(1/2 \leq d_0 < 1)$. Compared with Fig. 1, we observe that the difference in the standard errors of these two estimators becomes smaller while the difference in the biases remains more or less the same. Although $\widehat{d}_{LP}$ is still a better estimator than $\widehat{d}$ in this case, the advantage of $\widehat{d}_{LP}$ has clearly diminished with the increase in d_0.

Figure 4 provides results for the case $d_0 = 0.85$ and $\sigma_u^2 = 8$. Fig. 4(a) shows that the bias reduction from using $\widehat{d}$ is substantial. For example, when $m = 40$, the bias of $\widehat{d}_{LP}$ is -0.18, while that of $\widehat{d}$ is only -0.02. The evidence therefore suggests that $\widehat{d}$ is effective in reducing bias not only in stationary fractional component models but also in nonstationary models. Fig. 4(b) shows that the standard error of $\widehat{d}$ is only slightly larger than that of $\widehat{d}_{LP}$. The large bias reduction and small variance inflation lead to a smaller RMSE for $\widehat{d}$ over a wide range of m values, as shown in Fig. 4(c). In addition, the coverage probability based on $\widehat{d}_{LP}$ decreases very rapidly as m increases, whereas that based on $\widehat{d}$ decreases much more slowly. In fact, the coverage probability based on $\widehat{d}$ is close to 0.9 over a wide range of m values. Fig. 4(e) shows that the superior performance of the coverage probability of $\widehat{d}$ comes at the expense of having longer confidence intervals on average than those based on $\widehat{d}_{LP}$.

The simulations also reveal that the bias of $\widehat{d}_{LP}$ is always negative when $\sigma_u > 0$ and that the absolute value of the bias increases with σ_u^2, due to stronger contamination from the perturbations that this produces. In addition, $\widehat{d}$ is more effective in bias reduction for larger values of d_0. Intuitively, when d_0 is small, the bias of $\widehat{d}_{LP}$ is small no matter what value σ_u may take. For a large value of σ_u, the perturbation component dominates the fractional component, so that $\widehat{d}_{LP}$ would be around 0. In this case, the bias of $\widehat{d}_{LP}$ is small only because the true value of d_0 itself is small. Also, for small values of σ_u, the bias from contamination is naturally going to be small. Therefore, in both cases, the bias of $\widehat{d}_{LP}$ will be small when d_0 is small and there is not much scope for $\widehat{d}$ to manifest its bias-reducing capacity.

Finally, for the large sample size $n = 2048$, the qualitative comparisons made and conclusions reached for the $n = 512$ sample size continue to apply. Fig. 5 presents the results for one particular specification ($d_0 = 0.85, \sigma_u^2 = 8$) which shows that $\widehat{d}$ has a much smaller bias and a slightly larger variance than $\widehat{d}_{LP}$. The RMSE of $\widehat{d}$ is much smaller than that of $\widehat{d}_{LP}$ over a wide range of m values.

To sum up, the simulations show that, for fractional component processes, the new estimator $\widehat{d}$ has a lower bias, a higher standard deviation, and a lower RMSE in comparison to the LP estimator $\widehat{d}_{LP}$, corroborating the asymptotic theory. The lower bias generally leads to improved coverage probability in confidence intervals based on $\widehat{d}$ over a wide range of m. On the other hand, the lower standard deviation of $\widehat{d}_{LP}$ leads to shorter confidence intervals than those based on $\widehat{d}$.

6 Conclusion

In empirical applications it has become customary practice to investigate the order of integration of the variables in a model when nonstationarity is suspected. This practice is now being extended to include analyses of the degree of persistence using fractional models and estimates of long memory parameters. Nonetheless, for many time series, and particularly macroeconomic variables for which there is limited data, the actual degree of persistence in the data continues to be a controversial issue. The empirical resolution of this problem inevitably relies on our capacity to separate low-frequency behavior from high-frequency fluctuations and this is particularly difficult when short run fluctuations have high variance. Actual empirical results often depend critically on the discriminatory power of the statistical techniques being employed to implement the separation.

The model used in the present paper provides some assistance in this regard. It allows for an explicit components structure in which there are different sources and types of variation, thereby accommodating a separation of short and long memory components and allowing for fractional processes that are perturbed by weakly dependent effects. Compared to the conventional formulation of a pure fractional process like (2), perturbed fractional processes allow for multiple sources of high-frequency variation and, in doing so, seem to provide a richer setting for uncovering latent persistence in an observed time series. In particular, the model provides a mechanism for simultaneously capturing the effects of persistent and temporary shocks and seems realistic in economic and financial applications when there are many different sources of variation in the data. The new econometric methods we have introduced for estimating the fractional parameter in such models take account of the presence of additive disturbances, and help to achieve bias reduction and attain a faster rate of convergence. The asymptotic theory is easy to use and seems to work reasonably well in finite samples.

The methods of the paper can be extended in a number of directions. First, it is of interest to study the performance of the methods here under non-Gaussian errors, as in Deo and Hurvich (1999) for LP regression. Second, the nonlinear approximation approach can be used in combination with other estimators, such as the local Whittle estimator (Robinson 1995b), which seems natural in the present context because the procedure already uses optimization methods. In addition, the idea of using a nonlinear approximation can be applied to nonstationary fractional component models and used to adapt the methods which have been suggested elsewhere (e.g., Phillips, 1999, Shimotsu and Phillips, 2001) for estimating the memory parameter in such models to cases where there are fractional components.

Appendix of Proofs

Proof of Lemma 1. A spectral density satisfying Assumptions 1 and 2 also satisfies Assumptions 1 and 2 of Robinson (1995a). In consequence, the lemma follows from Theorem 2 of Robinson (1995a). Since we normalize the discrete Fourier transform by the spectral density $f_z^{1/2}(\lambda)$ instead of the power function $C_g^{-1/2}\lambda^{-d}$, (4.2) of Robinson (1995a) is always zero and the extra term $(\frac{j}{n})^{\min(\alpha,\beta)}$ in Robinson (1995a) does not arise in our case. □

Proof of Lemma 3. Note that

$$\frac{1}{m}\sum_{j=1}^{m} c_j U_j = \frac{1}{m}\sum_{j=1}^{[\log^2 m]} c_j U_j + \frac{1}{m}\sum_{j=[\log^2 m]+1}^{m} c_j U_j \equiv F_1 + F_2. \tag{A.1}$$

But $E\sup_{(d,\beta)\in\Theta}|F_1|$ is less than

$$E\frac{1}{m}\sum_{j=1}^{[\log^2 m]} \sup_{(d,\beta)\in\Theta}|c_j||U_j| \leqslant \frac{\log^p m}{m}\sum_{j=1}^{[\log^2 m]}(EU_j^2)^{1/2} = O(\log^{p+2} m/m) \tag{A.2}$$

by Lemma 2(b). Hence

$$\sup_{(d,\beta)\in\Theta}|F_1| = O_p(\log^{p+2} m/m) = O_p(\frac{\log^p m}{\sqrt{m}}). \tag{A.3}$$

Let

$$s_r = \sum_{k=[\log^2 m]+1}^{r} U_r, r = [\log^2 m]+1, ..., m \text{ and } s_{[\log^2 m]} = 0. \tag{A.4}$$

Then, from Lemma 2(a), (c) and (d), it follows that

$$\begin{aligned} Es_r^2 &= \sum_{k=[\log^2 m]+1}^{r} EU_k^2 + 2\sum_{[\log^2 m]+1\leq k<j\leq r} EU_j U_k \\ &= \sum_{k=[\log^2 m]+1}^{r} (\frac{\pi^2}{6} + k^{-1}\log k) + 2\sum_{[\log^2 m+1]\leq k<j<r}^{r} O(k^{-2}\log^2 j) \qquad \text{(A.5)} \\ &= O(r) + O(r\log^2 r/\log^2 m), \end{aligned}$$

which implies $s_r = O_p(r^{1/2})$. Using this result and partial summation, we have:

$$\begin{aligned}
\sup_{(d,\beta)\in\Theta} |F_2| &\leq \sup_{(d,\beta)\in\Theta} \left| \frac{1}{m} \sum_{j=\log^2 m+1}^{m} c_j U_j \right| \\
&= \sup_{(d,\beta)\in\Theta} \frac{1}{m} \left| \sum_{j=\log^2 m+1}^{m} s_{j-1}(c_{j-1} - c_j) \right| + \sup_{(d,\beta)\in\Theta} \frac{1}{m} |s_m c_m| \\
&= \frac{1}{m} \sum_{j=\log^2 m+1}^{m} O_p(j^{1/2}) O(j^{-1} \log^q m) + O_p(\frac{\log^p m}{\sqrt{m}}) \\
&= \frac{\log^q m}{m} \sum_{j=\log^2 m+1}^{m} O_p(j^{-1/2}) + O_p(\frac{\log^p m}{\sqrt{m}}) \\
&= O_p(\frac{\log^q m}{\sqrt{m}}) + O_p(\frac{\log^p m}{\sqrt{m}}) = O_p(\frac{\log^{\max(p,q)} m}{\sqrt{m}}). \qquad (A.6)
\end{aligned}$$

Combining (A.3) with (A.6), we get $\sup_{(d,\beta)\in\Theta} \left| \frac{1}{m} \sum_{j=1}^{m} c_j U_j \right| = O_p(\frac{\log^{\max(p,q)} m}{\sqrt{m}})$.
□

Proof of Theorem 1. When $\sigma_u = 0$, the theorem is essentially the same as results already established in HDB. Only one modification is needed. HDB use $-2\log(2\sin\lambda_j/2)$ as one of the regressors while we employ $-2\log\lambda_j$. The use of $-2\log\lambda_j$ rather than $-2\log(2\sin\lambda_j/2)$ has no effect on the asymptotic variance, but it does affect the asymptotic bias. This is because the asymptotic bias comes from the dominating term in ε_j and this term is different for different regressors. Using $-2\log(2\sin\lambda_j/2)$ as the regressor yields

$$\varepsilon_j = \log f_w(\lambda_j) - \log f_w(0) = \left(\frac{f_w''(0)}{2f_w'(0)} \right) \lambda_j^2 (1 + o(1)). \qquad (A.7)$$

In contrast, using $-2\log\lambda_j$ as the regressor yields

$$\begin{aligned}
\varepsilon_j &= \log f_w(\lambda_j) - \log f_w(0) - 2d_0 \left(\log(2\sin\frac{\lambda_j}{2}) - \log\lambda_j \right) \\
&= \left(\frac{f_w''(0)}{2f_w'(0)} + \frac{d_0}{12} \right) \lambda_j^2 (1 + o(1)). \qquad (A.8)
\end{aligned}$$

With this adjustment, the arguments in HDB go through without further change.

Now consider the case $\sigma_u \neq 0$. Rewrite the spectral density of z_t as

$$f_z(\lambda) = \lambda^{-2d_0} g(\lambda), \qquad (A.9)$$

where $g(\lambda) = (\lambda^{-1} 2\sin\lambda/2)^{-2d_0} f^*(\lambda)$. Since

$$g(\lambda) - g(0) = (1 + O(\lambda^2)) \left(f_w(0) + \lambda^{2d_0} f_u(0) + O(\lambda^2) \right) - f_w(0) = O(\lambda^{2d_0}) \quad (A.10)$$

as $\lambda \to 0+$, $g(\lambda)$ is smooth of order $2d_0$. Combining this with our assumption that $m \to \infty$ and $m^{4d_0+1}/n^{4d_0} = O(1)$ verifies Assumptions 1 and 2 of Andrews and Guggenberger (1999). Hence their Theorem 1 is valid with $r = 0$, $s = 2d_0$ and $q = 2d_0$. It is easy to show that the term $O(m^q/n^q)$ in their theorem is actually $-(2\pi)^{2d_0} \frac{f_u(0)}{f_w(0)} \frac{d_0}{(2d_0+1)^2} \frac{m^{2d_0}}{n^{2d_0}}$. Andrews and Guggenberger established asymptotic normality under their assumption 3 that $m^{4d_0+1}/n^{4d_0} = o(1)$. In fact, asymptotic normality holds under our assumption $m^{4d_0+1}/n^{4d_0} = O(1)$ as long as an asymptotic bias of order $O(1)$ is allowed.□

Proof of Theorem 2. Let $V_j(d,\beta) = 2(d-d_0)\log\lambda_j - \beta\lambda_j^{2d} + \beta_0\lambda_j^{2d_0}$. Then we can decompose $m^{-1}Q(d,\beta)$ into three parts as follows:

$$\frac{1}{m}Q(d,\beta) = \frac{1}{m}\sum_{j=1}^{m}\{U_j - \overline{U} + \varepsilon_j - \overline{\varepsilon} + V_j - \overline{V}\}^2 \tag{A.11}$$

$$= \frac{1}{m}\sum_{j=1}^{m}(V_j - \bar{V})^2 + \frac{2}{m}\sum_{j=1}^{m}(U_j + \varepsilon_j)(V_j - \bar{V}) + \frac{1}{m}\sum_{j=1}^{m}(U_j + \varepsilon_j - \bar{U} - \bar{\varepsilon})^2,$$

where the dependence on (d,β) has been suppressed for notational simplicity. Since $\frac{1}{m}\sum_{j=1}^{m}(U_j + \varepsilon_j - \bar{U} - \bar{\varepsilon})^2$ is independent of (d,β), we only need to consider the first two terms.

Part (a) We prove part (a) by showing that $\frac{2}{m}\sum_{j=1}^{m}(U_j + \varepsilon_j)(V_j - \bar{V}) = o_p(1)$ uniformly in (d,β) and $\frac{1}{m}\sum_{j=1}^{m}(V_j - \bar{V})^2$ converges uniformly to a function, which has a unique minimizer d_0.

First, we show

$$\sup_{(d,\beta)\in\Theta}\left|\frac{1}{m}\sum_{j=1}^{m}U_j(V_j - \bar{V})\right| = O_p(\frac{1}{\sqrt{m}}). \tag{A.12}$$

We proceed by verifying the conditions in Lemma 3. The first condition holds because

$$\begin{aligned}
&\sup_{(d,\beta)\in\Theta}|V_j(d,\beta) - \bar{V}(d,\beta)| \\
\leq\ & 2\sup_{(d,\beta)\in\Theta}|d-d_0|\ |\log\lambda_j - \frac{1}{m}\sum_{j=1}^{m}\log\lambda_j| + 2\sup_{(d,\beta)\in\Theta}|\beta||\lambda_j^{2d} - \frac{1}{m}\sum_{j=1}^{m}\lambda_j^{2d}| \\
=\ & 2\sup_{(d,\beta)\in\Theta}|d-d_0|\log m + O(1) = O(\log m) \text{ uniformly over } j. \qquad (A.13)
\end{aligned}$$

The second condition holds because

$$\begin{aligned}
&\sup_{(d,\beta)\in\Theta}|V_j(d,\beta) - V_{j-1}(d,\beta)| \\
\leq\ & 2\sup_{(d,\beta)\in\Theta}|d-d_0|\ |\log(1-\frac{1}{j})| + 2\sup_{(d,\beta)\in\Theta}|\beta\lambda_j^{2d}(1-(1-\frac{1}{j})^{2d})| \\
=\ & O(\frac{1}{j}) \text{ for all } j, \qquad (A.14)
\end{aligned}$$

where the final line follows from the fact that $\sup_{(d,\beta)\in\Theta}|1-(1-\frac{1}{j})^{2d}| = O(\frac{1}{j})$. In addition,

$$\begin{aligned}
&|V_m(d,\beta)-\bar{V}(d,\beta)| \\
=\;& 2\sup_{(d,\beta)\in\Theta}|d-d_0||\log\lambda_m-\frac{1}{m}\sum_{j=1}^{m}\log\lambda_m|+O(1) \\
=\;& 2\sup_{(d,\beta)\in\Theta}|d-d_0||\log m-\frac{1}{m}\sum_{j=1}^{m}\log j|+O(1)=O(1). \qquad \text{(A.15)}
\end{aligned}$$

Hence (A.12) is satisfied.

Next, we show that $\sup_{(d,\beta)\in\Theta}|\frac{1}{m}\sum_{j=1}^{m}\varepsilon_j(V_j(d,\beta)-\bar{V}(d,\beta))| = O_p(\lambda_m^{4d_0})$. Under Assumptions 1 and 2,

$$\begin{aligned}
\varepsilon_j &= \log f^*(\lambda_j)-\log f_w(0)-(2\sin\frac{\lambda_j}{2})^{2d_0}\beta_0-2d_0[\log(2\sin\frac{\lambda_j}{2})-\log\lambda_j] \\
&= O(\lambda_j^r)=O(\lambda_j^{4d_0}), \qquad \text{(A.16)}
\end{aligned}$$

so we have, using (A.14) and (A.15)

$$\begin{aligned}
&\sup_{(d,\beta)\in\Theta}|\frac{1}{m}\sum_{j=1}^{m}\varepsilon_j(V_j(d,\beta)-\bar{V}(d,\beta))| \\
\leq\;& \sup_{(d,\beta)\in\Theta}\frac{1}{m}\left|\sum_{j=1}^{m}\sum_{r=1}^{j-1}\varepsilon_r(V_{j-1}(d,\beta)-V_j(d,\beta))\right|+\sup_{(d,\beta)\in\Theta}\frac{1}{m}|\sum_{j=1}^{m}\varepsilon_j||V_m(d,\beta)-\bar{V}(d,\beta)| \\
=\;& \lambda_m^{4d_0}\frac{1}{m}\left|\sum_{j=1}^{m}\sum_{r=1}^{j-1}O_p(\frac{r}{m})^{4d_0}(\frac{1}{j})\right|+O_p(\lambda_m^{4d_0})=O_p(\lambda_m^{4d_0})=o_p(1). \qquad \text{(A.17)}
\end{aligned}$$

Finally,

$$\begin{aligned}
\frac{1}{m}\sum_{j=1}^{m}(V_j-\bar{V})^2 &= \frac{1}{m}\sum_{j=1}^{m}\left(2(d-d_0)(\log(\frac{j}{m})-\frac{1}{m}\sum_{k=1}^{m}\log(\frac{k}{m}))+o(1)\right)^2 \\
&= 4(d-d_0)^2\left(\frac{1}{m}\sum_{j=1}^{m}\log^2(\frac{j}{m})-(\frac{1}{m}\sum_{k=1}^{m}\log(\frac{k}{m}))^2\right)+o(1) \\
&= 4(d-d_0)^2(1+o(1)), \qquad \text{(A.18)}
\end{aligned}$$

where $o(\cdot)$ holds uniformly over $(d,\beta)\in\Theta$. Here we have employed $\frac{1}{m}\sum_{j=1}^{m}\log^2(\frac{j}{m})-\left(\frac{1}{m}\sum_{k=1}^{m}\log(\frac{k}{m})\right)^2 = 1+o(1)$.

Let $D_\delta = \{d : |d - d_0| > \delta\}$. In view of (A.12), (A.17) and (A.18), we have

$$\begin{aligned}
P\left(\widehat{d} \in D_\delta\right) &\leq P(Q(\widehat{d}, \widehat{\beta}) \leq Q(d_0, \beta_0)) \\
&\leq P\left(\min_{d \in D_\delta, \beta \in B} \left[\frac{1}{m}\sum_{j=1}^{m}(V_j - \bar{V})^2 + \frac{2}{m}\sum_{j=1}^{m}(U_j + \varepsilon_j)(V_j - \bar{V})\right] \leq 0\right) \\
&\leq P\left(\min_{d \in D_\delta, \beta \in B} \frac{1}{m}\sum_{j=1}^{m}(V_j - \bar{V})^2 \leq \sup_{d \in D_\delta, \beta \in B}\left|\frac{2}{m}\sum_{j=1}^{m}(U_j + \varepsilon_j)(V_j - \bar{V})\right|\right) \\
&\leq P\left(\min_{d \in D_\delta, \beta \in \beta} \frac{1}{m}\sum_{j=1}^{m}(V_j - \bar{V})^2 \leq o(1)\right) \\
&\leq P(\min_{d \in D_\delta} 4(d - d_0)^2 \leq o(1)) \to 0, \qquad \text{(A.19)}
\end{aligned}$$

which completes the proof.

Part (b) Compared with $\log \lambda_j$, $\lambda_j^{2d_0}$ is negligible since $d_0 > 0$. Due to the difference in the orders of magnitude of the regressors, it is not straightforward to establish the consistency of $\widehat{\beta}$. In fact, we proceed by showing first that $\widehat{d}$ converges to d_0 at some preliminary rate and then go on to show that $\widehat{d} - d_0 = O_p((\frac{m}{n})^{2d_0})$. We obtain this rate sequentially.

First, we show that $\widehat{d} - d_0 = o_p((\frac{m}{n})^{d_1/2})$. From $Q(\widehat{d}, \widehat{\beta}) - Q(d_0, \beta_0) \leq 0$, we get

$$\begin{aligned}
&\frac{1}{m}\sum_{j=1}^{m}(V_j(\widehat{d}, \widehat{\beta}) - \bar{V}(\widehat{d}, \widehat{\beta}))^2 \\
\leq\ & -\frac{2}{m}\sum_{j=1}^{m}(U_j + \varepsilon_j)(V_j(\widehat{d}, \widehat{\beta}) - \bar{V}(\widehat{d}, \widehat{\beta})) \qquad \text{(A.20)} \\
\leq\ & \sup_{(d,\beta) \in \Theta}\left|\frac{2}{m}\sum_{j=1}^{m}(U_j + \varepsilon_j)(V_j(d, \beta) - \bar{V}(d, \beta))\right| \\
=\ & O_p(\frac{1}{\sqrt{m}}) + O_p(\lambda_m^{4d_0}) = o_p\left((\frac{m}{n})^{2d_1}\right), \qquad \text{(A.21)}
\end{aligned}$$

where the last equality follows from the assumptions that $n^{4d_0(1+\Delta)}/m^{4d_0(1+\Delta)+1} = o(1)$ and that $d \geq d_1 > 0$. But $\frac{1}{m}\sum_{j=1}^{m}(V_j(\widehat{d}, \widehat{\beta}) - \bar{V}(\widehat{d}, \widehat{\beta}))^2$ equals

$$\begin{aligned}
&\frac{1}{m}\sum_{j=1}^{m}\left(2(\widehat{d} - d_0)(\log(\frac{j}{m}) - \frac{1}{m}\sum_{j=1}^{m}\log(\frac{j}{m})) + O(\lambda_m^{2\widehat{d}}) + O(\lambda_m^{2d_0})\right)^2 \\
=\ & 4(\widehat{d} - d_0)^2(1 + o(1)) + O(\lambda_m^{2d_0}) + O(\lambda_m^{2\widehat{d}}) \\
=\ & 4(\widehat{d} - d_0)^2(1 + o(1)) + O(\frac{m}{n})^{2d_1}. \qquad \text{(A.22)}
\end{aligned}$$

Therefore,

$$4(\widehat{d} - d_0)^2(1 + o(1)) + O_p\left((\frac{m}{n})^{2d_1}\right) \leq o_p\left((\frac{m}{n})^{2d_1}\right), \qquad \text{(A.23)}$$

which implies that $\widehat{d} - d_0$ is at most $O_p((\frac{m}{n})^{d_1})$. Thus $\widehat{d} - d_0 = o_p((\frac{m}{n})^{d_1/2})$.

Second, we show that $\widehat{d} - d_0 = o_p\left((\frac{m}{n})^{d_0(1+\Delta)}\right)$. Since $\widehat{d} - d_0 = o_p((\frac{m}{n})^{d_1/2})$, we only need consider $d \in D_n' = \{d : |d - d_0| < \varepsilon(\frac{m}{n})^{d_1/2}\}$ for some small $\varepsilon > 0$. Approximating sums by integrals and using the formulae:

$$\frac{1}{m}\sum_{j=1}^{m}(\frac{j}{m})^k \log(\frac{j}{m}) = -\frac{1}{(k+1)^2} + o(1), k \geq 0, \tag{A.24}$$

and

$$\frac{1}{m}\sum_{j=1}^{m}(\frac{j}{m})^k \log^2(\frac{j}{m}) = \frac{2}{(k+1)^3} + o(1), k \geq 0, \tag{A.25}$$

we deduce that $\frac{1}{m}\sum_{j=1}^{m} V_j^2(d,\beta) - \left(\bar{V}(d,\beta)\right)^2$ is equal to

$$\begin{aligned}
& \left(8(d-d_0)^2 + \frac{\beta^2\lambda_m^{4d}}{4d+1} + \frac{\beta_0^2\lambda_m^{4d_0}}{4d_0+1} + \frac{4(d-d_0)\beta\lambda_m^{2d}}{2d+1} - \frac{4(d-d_0)\beta_0\lambda_m^{2d_0}}{2d_0+1}\right)(1+o(1)) \\
& -\frac{2\beta\beta_0\lambda_m^{2d+2d_0}}{2d+2d_0+1}(1+o(1)) - \left(2(d-d_0) + \frac{\beta\lambda_m^{2d}}{2d+1} - \frac{\beta_0\lambda_m^{2d_0}}{2d_0+1}\right)^2(1+o(1)) \\
= & \left(4(d-d_0)^2 + \left[\frac{2d\beta\lambda_m^{2d}}{(2d+1)\sqrt{4d+1}} - \frac{2d_0\beta_0\lambda_m^{2d_0}}{(2d_0+1)\sqrt{4d_0+1}}\right]^2\right)(1+o(1)) \qquad \text{(A.26)} \\
& + \frac{8dd_0\beta\beta_0\lambda_m^{2d+2d_0}}{(2d+1)(2d_0+1)}\left(\frac{1}{\sqrt{(4d+1)(4d_0+1)}} - \frac{1}{2d+2d_0+1}\right)(1+o(1)) \\
= & \; 4(d-d_0)^2 + O(\lambda_m^{4d_0}) = 4(d-d_0)^2 + o(\lambda_m^{2d_0(1+\Delta)}), \qquad \text{(A.27)}
\end{aligned}$$

where the $O(\cdot)$ and $o(\cdot)$ terms hold uniformly over $(d,\beta) \in D_n' \times B$. The last line follows because when $|d - d_0| \leq \varepsilon(\frac{m}{n})^{d_1/2}$,

$$\begin{aligned}
\lambda_m^{2d} &= \lambda_m^{2d_0}\lambda_m^{2d-2d_0} = \lambda_m^{2d_0}\exp((2d-2d_0)\log\lambda_m) \\
&\leq const\lambda_m^{2d_0}\exp(\lambda_m^{d_1/2}|\log\lambda_m|) = O(\lambda_m^{2d_0}).
\end{aligned} \tag{A.28}$$

Using $Q(\widehat{d},\widehat{\beta}) - Q(d_0,\beta_0) \leq 0$ again, we have

$$\frac{1}{m}\sum_{j=1}^{m}(V_j(\widehat{d},\widehat{\beta}) - \bar{V}(\widehat{d},\widehat{\beta}))^2 \leq O_p(\frac{1}{\sqrt{m}}) + O_p(\lambda_m^{4d_0}) = o_p\left((\frac{m}{n})^{2d_0(1+\Delta)}\right), \tag{A.29}$$

where the equality follows from the assumption $n^{4d_0(1+\Delta)}/m^{4d_0(1+\Delta)+1} = o(1)$. Combining (A.27) and (A.29), we get

$$4(\widehat{d} - d_0)^2 + o(\lambda_m^{2d_0(1+\Delta)}) \leq o_p\left((\frac{m}{n})^{2d_0(1+\Delta)}\right). \tag{A.30}$$

Hence $\widehat{d} - d_0 = o_p((\frac{m}{n})^{d_0(1+\Delta)})$.

Next, we show that $\widehat{d}-d_0 = o_p\left((\frac{m}{n})^{3d_0(1+\Delta)/2}\right)$. We first prove that $\left|\frac{1}{m}\sum_{j=1}^m U_j(V_j-\bar{V})\right| = o_p\left((\frac{m}{n})^{3d_0(1+\Delta)}\right)$ and $\left|\frac{1}{m}\sum_{j=1}^m \varepsilon_j(V_j-\bar{V})\right| = o_p\left((\frac{m}{n})^{3d_0(1+\Delta)}\right)$ uniformly in $(d,\beta)\in D_n''\times B$, where $D_n''=\{d:|d-d_0|<\varepsilon(\frac{m}{n})^{d_0(1+\Delta)}\}$.

Following the same steps as in the proof of Lemma 3, we compute the orders of $|V_j(d,\beta)-\bar{V}(d,\beta)|$, $|V_m(d,\beta)-\bar{V}(d,\beta)|$ and $|V_j(d,\beta)-V_{j-1}(d,\beta)|$ as follows.

First, $\sup_{(d,\beta)\in D_n''\times B}|V_j(d,\beta)-\bar{V}(d,\beta)|$ is bounded by

$$\begin{aligned}
& 2\sup_{(d,\beta)\in D_n''\times B}|d-d_0||\log\lambda_j-\frac{1}{m}\sum_{j=1}^m\log\lambda_j| + 2\sup_{(d,\beta)\in D_n''\times B}|\beta||\lambda_j^{2d}-\frac{1}{m}\sum_{j=1}^m\lambda_j^{2d}| \\
= & 2\sup_{(d,\beta)\in\Theta}|d-d_0|\log m + O(\lambda_m^{2d_0}) \\
= & O((\frac{m}{n})^{d_0(1+\Delta)}\log m) \text{ uniformly over } j. \qquad \text{(A.31)}
\end{aligned}$$

Similarly, $\sup_{(d,\beta)\in D_n''\times B}|V_m(d,\beta)-\bar{V}(d,\beta)|$ is bounded by

$$\begin{aligned}
& 2\sup_{(d,\beta)\in D_n''\times B}|d-d_0||\log\lambda_m-\frac{1}{m}\sum_{j=1}^m\log\lambda_m| + O(\lambda_m^{2d_0}) \\
= & 2\sup_{(d,\beta)\in D_n''\times B}|d-d_0||\log m-\frac{1}{m}\sum_{j=1}^m\log j| + O(\lambda_m^{2d_0}) \\
= & O\left((\frac{m}{n})^{d_0(1+\Delta)}\right). \qquad \text{(A.32)}
\end{aligned}$$

Furthermore, $\sup_{(d,\beta)\in D_n''\times B}|V_j(d,\beta)-V_{j-1}(d,\beta)|$ is not greater than

$$\begin{aligned}
& 2\sup_{(d,\beta)\in D_n''\times B}||d-d_0||\log(1-\frac{1}{j})| + 2\sup_{(d,\beta)\in D_n''\times B}|\beta\lambda_j^{2d}(1-(1-\frac{1}{j})^{2d})| \\
= & O(\frac{1}{j}(\frac{m}{n})^{d_0(1+\Delta)}) \text{ for all } j. \qquad \text{(A.33)}
\end{aligned}$$

Invoking the same argument as in the proof of Lemma (3), we obtain

$$\begin{aligned}
\sup_{(d,\beta)\in D_n''\times B}\left|\frac{1}{m}\sum_{j=1}^m U_j(V_j-\bar{V})\right| &= O_p\left((\frac{m}{n})^{d_0(1+\Delta)}\frac{1}{\sqrt{m}}\right) = o_p\left((\frac{m}{n})^{3d_0(1+\Delta)}\right), \text{ and} \\
\sup_{(d,\beta)\in D_n''\times B}\left|\frac{1}{m}\sum_{j=1}^m \varepsilon_j(V_j-\bar{V})\right| &= o_p\left((\frac{m}{n})^{d_0(1+\Delta)+4d_0}\right) = o_p\left((\frac{m}{n})^{3d_0(1+\Delta)}\right), \qquad \text{(A.34)}
\end{aligned}$$

as desired.

In addition, it follows from (A.27) that when $d\in D_n''$, $\frac{1}{m}\sum_{j=1}^m(V_j-\bar{V})^2 = 4(d-d_0)^2(1+o(1))+o(\lambda_m^{3d_0(1+\Delta)})$. Applying the same argument as before, we get

$$4(\widehat{d}-d_0)^2(1+o(1))+o(\lambda_m^{3d_0(1+\Delta)}) \le o_p(\frac{m}{n})^{3d_0(1+\Delta)}, \qquad \text{(A.35)}$$

and so $\widehat{d}-d_0=o_p((\frac{m}{n})^{3d_0(1+\Delta)/2})$.

Repeating the procedure again we obtain $\widehat{d}-d_0 = o_p\left((\frac{m}{n})^{7d_0(1+\Delta)/4}\right)$ if $7(1+\Delta)/4<2$. Further iterations of this procedure lead to $\widehat{d}-d_0=o_p\left((\frac{m}{n})^{(2-2^{-k})(1+\Delta)}\right)$, $k=0,1,2,3,...$ if $(2-2^{-k})(1+\Delta)<2$. We stop the iteration if we obtain $\widehat{d}-d_0=o_p\left((\frac{m}{n})^{(2-2^{-k_0})(1+\Delta)}\right)$ for some $k_0\geq 0$ such that $(2-2^{-k_0})(1+\Delta)<2$ and $(4-2^{-k_0})(1+\Delta)\geq 4$. In this case, we have

$$\begin{aligned}
&\left|\frac{1}{m}\sup_{(d,\beta)\in D_n^*\times B}\sum_{j=1}^{m}U_j(V_j-\bar{V})\right| \\
&= O_p\left((\frac{m}{n})^{(2-2^{-k_0})(1+\Delta)d_0}\frac{1}{\sqrt{m}}\right) \\
&= o_p\left((\frac{m}{n})^{(4-2^{-k_0})(1+\Delta)d_0}\right)=o_p\left((\frac{m}{n})^{4d_0}\right), \qquad (A.36)
\end{aligned}$$

and

$$\sup_{(d,\beta)\in D_n^*\times B}\left|\frac{1}{m}\sum_{j=1}^{m}\varepsilon_j(V_j-\bar{V})\right|=o_p\left((\frac{m}{n})^{4d_0}\right), \qquad (A.37)$$

where $D_n^*=\{d:|d-d_0|<\varepsilon(\frac{m}{n})^{(2-2^{-k_0})(1+\Delta)}\}$. Applying the same argument as before, we deduce

$$4(\widehat{d}-d_0)^2(1+o(1))+O(\lambda_m^{4d_0})\leq o_p(\frac{m}{n})^{4d_0}. \qquad (A.38)$$

In consequence, $\widehat{d}-d_0=O_p(\frac{m}{n})^{2d_0}$.

Now, since $(2d+2d_0+1)^2-(4d+1)(4d_0+1)=4d^2-8dd_0+4d_0^2=4(d-d_0)^2>0$, we deduce from (A.26) that

$$\frac{1}{m}\sum_{j=1}^{m}(V_j-\bar{V})^2\geq\left(\frac{2d\beta\lambda_m^{2d}}{(2d+1)\sqrt{4d+1}}-\frac{2d_0\beta_0\lambda_m^{2d_0}}{(2d_0+1)\sqrt{4d_0+1}}\right)^2(1+o(1)). \qquad (A.39)$$

In view of $\frac{1}{m}\sum_{j=1}^{m}(V_j(\widehat{d},\widehat{\beta})-\bar{V}(\widehat{d},\widehat{\beta}))^2\leq o_p(\lambda_m^{4d_0})$, we obtain

$$\left(\frac{2\widehat{d}\widehat{\beta}\lambda_m^{2\widehat{d}}}{(2\widehat{d}+1)\sqrt{4\widehat{d}+1}}-\frac{2d_0\beta_0\lambda_m^{2d_0}}{(2d_0+1)\sqrt{4d_0+1}}\right)^2(1+o(1))\leq o_p(\lambda_m^{4d_0}). \qquad (A.40)$$

But

$$\frac{2\widehat{d}\widehat{\beta}\lambda_m^{2\widehat{d}}}{(2\widehat{d}+1)\sqrt{4\widehat{d}+1}} - \frac{2d_0\beta_0\lambda_m^{2d_0}}{(2d_0+1)\sqrt{4d_0+1}}$$

$$= \frac{2\widehat{d}\widehat{\beta}\lambda_m^{2\widehat{d}}}{(2\widehat{d}+1)\sqrt{4\widehat{d}+1}} - \frac{2\widehat{d}\beta_0\lambda_m^{2\widehat{d}}}{(2\widehat{d}+1)\sqrt{4\widehat{d}+1}} + \frac{2\widehat{d}\beta_0\lambda_m^{2\widehat{d}}}{(2\widehat{d}+1)\sqrt{4\widehat{d}+1}} - \frac{2\widehat{d}\beta_0\lambda_m^{2d_0}}{(2\widehat{d}+1)\sqrt{4\widehat{d}+1}}$$

$$+\frac{2\widehat{d}\beta_0\lambda_m^{2d_0}}{(2\widehat{d}+1)\sqrt{4\widehat{d}+1}} - \frac{2d_0\beta_0\lambda_m^{2d_0}}{(2d_0+1)\sqrt{4d_0+1}}$$

$$= \frac{2\widehat{d}\lambda_m^{2\widehat{d}}\left(\widehat{\beta}-\beta_0\right)}{(2\widehat{d}+1)\sqrt{4\widehat{d}+1}} + \frac{2\widehat{d}\beta_0(\lambda_m^{2\widehat{d}}-\lambda_m^{2d_0})}{(2\widehat{d}+1)\sqrt{4\widehat{d}+1}}$$

$$+\left(\frac{2\widehat{d}}{(2\widehat{d}+1)\sqrt{4\widehat{d}+1}} - \frac{2d_0}{(2d_0+1)\sqrt{4d_0+1}}\right)O(\lambda_m^{2d_0})$$

$$= \frac{2\widehat{d}\lambda_m^{2\widehat{d}}}{(2\widehat{d}+1)\sqrt{4\widehat{d}+1}}\left(\widehat{\beta}-\beta_0\right) + O_p(\left|\lambda_m^{2\widetilde{d}}\log\lambda_m(\widehat{d}-d_0)\right|) + O_p(\lambda_m^{4d_0})$$

$$= \frac{2\widehat{d}\lambda_m^{2\widehat{d}}}{(2\widehat{d}+1)\sqrt{4\widehat{d}+1}}\left(\widehat{\beta}-\beta_0\right) + o_p(\lambda_m^{3d_0}), \tag{A.41}$$

where $\widetilde{d}$ is between $\widehat{d}$ and d_0. So

$$\left(\frac{2\widehat{d}\lambda_m^{2\widehat{d}}}{(2\widehat{d}+1)\sqrt{4\widehat{d}+1}}\left(\widehat{\beta}-\beta_0\right) + o_p(\lambda_m^{3d_0})\right)^2 \leq o_p(\lambda_m^{4d_0}). \tag{A.42}$$

This implies that

$$\frac{4\widehat{d}\lambda_m^{4(\widehat{d}-d_0)}}{(2\widehat{d}+1)^2\left(4\widehat{d}+1\right)}(\widehat{\beta}-\beta_0)^2 \leq o_p(\lambda_m^{4d_0}), \tag{A.43}$$

from which we deduce that $\widehat{\beta}-\beta_0 = o_p(1)$. □

Proof of Lemma 5

Part (a) The (2,2) element of $\sup_{\theta\in\Theta_n} ||D_n^{-1}(H_n(d,\beta) - J_n(d,\beta))D_n^{-1}||$ is zero, so it suffices to consider the (1,1) and (1,2) elements. Since

$$I_{z_j} + 2d\log\lambda_j - \beta\lambda_j^{2d} = \alpha_0 + U_j + \varepsilon_j + (d-d_0)\log\lambda_j^2 + \beta_0\lambda_j^{2d_0} - \beta\lambda_j^{2d},$$

$\sup_{\theta\in\Theta_n} |\frac{\beta}{m}\sum_{j=1}^m e_j(\log\lambda_j^2)^2\lambda_j^{2d}|$, the (1,1) element, is bounded by $L_1+L_2+L_3+L_4$,

where

$$
\begin{aligned}
L_1 &= \sup_{\theta\in\Theta_n}\left|\frac{\beta}{m}\sum_{j=1}^{m}\left((\log\lambda_j^2)^2\lambda_j^{2d}-\frac{1}{m}\sum_{k=1}^{m}(\log\lambda_j^2)^2\lambda_j^{2d}\right)U_j\right|,\\
L_2 &= \sup_{\theta\in\Theta_n}\left|\frac{\beta}{m}\sum_{j=1}^{m}\left((\log\lambda_j^2)^2\lambda_j^{2d}-\frac{1}{m}\sum_{k=1}^{m}(\log\lambda_j^2)^2\lambda_j^{2d}\right)\varepsilon_j\right|,\\
L_3 &= \sup_{\theta\in\Theta_n}\left|\frac{\beta}{m}\sum_{j=1}^{m}\left((\log\lambda_j^2)^2\lambda_j^{2d}-\frac{1}{m}\sum_{k=1}^{m}(\log\lambda_j^2)^2\lambda_j^{2d}\right)(d-d_0)\log\lambda_j^2\right|,\ \text{and}\\
L_4 &= \sup_{\theta\in\Theta_n}\left|\frac{\beta}{m}\sum_{j=1}^{m}\left((\log\lambda_j^2)^2\lambda_j^{2d}-\frac{1}{m}\sum_{k=1}^{m}(\log\lambda_j^2)^2\lambda_j^{2d}\right)\left(\beta_0\lambda_j^{2d_0}-\beta\lambda_j^{2d}\right)\right|.
\end{aligned}
\tag{A.44}
$$

We first show that $L_1=o_p(1)$. Note that $\log^2(\lambda_j^2)\lambda_j^{2d}-\frac{1}{m}\sum_{k=1}^{m}\log^2(\lambda_k^2)\lambda_k^{2d}$ equals

$$
\begin{aligned}
&4\log^2\lambda_m\left(\lambda_j^{2d}-\frac{1}{m}\sum_{k=1}^{m}\lambda_k^{2d}\right)+8\log\lambda_m\left(\log(\frac{j}{m})\lambda_j^{2d}-\frac{1}{m}\sum_{k=1}^{m}\log(\frac{k}{m})\lambda_k^{2d}\right)\\
&+4\log^2(\frac{j}{m})\lambda_j^{2d}-\frac{4}{m}\sum_{k=1}^{m}\log^2(\frac{k}{m})\lambda_k^{2d}.
\end{aligned}
\tag{A.45}
$$

L_1 is thus bounded by $\sup_{\theta\in\Theta_n}|4\beta\lambda_m^{2d}|(\log^2\lambda_m L_{11}+2|\log\lambda_m|L_{12}+L_{13})$, where

$$
L_{1i+1}=\sup_{\theta\in\Theta_n}\left|\frac{1}{m}\sum_{j=1}^{m}\left((\frac{j}{m})^{2d}\log^i(\frac{j}{m})-\frac{1}{m}\sum_{k=1}^{m}(\frac{k}{m})^{2d}\log^i(\frac{k}{m})\right)U_j\right|,i=0,1,2.
\tag{A.46}
$$

It follows from Lemma 3 that $L_{1i+1}=O_p(\frac{\log^i m}{\sqrt{m}})$. The first condition is satisfied because

$$
\sup_{\theta\in\Theta_n}\left|(\frac{j}{m})^{2d}\log^i\left(\frac{j}{m}\right)-\frac{1}{m}\sum_{j=1}^{m}(\frac{j}{m})^{2d}\log^i\left(\frac{j}{m}\right)\right|=O(\log^i(m)).
\tag{A.47}
$$

The second condition is satisfied because

$$\begin{aligned}
& \sup_{\theta\in\Theta_n}\left|(\frac{j}{m})^{2d}\log^i\left(\frac{j}{m}\right)-(\frac{j-1}{m})^{2d}\log^i\left(\frac{j-1}{m}\right)\right| \\
\leq & \sup_{\theta\in\Theta_n}\left|(\frac{j}{m})^{2d}\log^i\left(\frac{j}{m}\right)-(\frac{j-1}{m})^{2d}\log^i\left(\frac{j}{m}\right)\right| \\
& +\left|(\frac{j-1}{m})^{2d}\log^i\left(\frac{j}{m}\right)-(\frac{j-1}{m})^{2d}\log^i\left(\frac{j-1}{m}\right)\right| \\
\leq & \sup_{\theta\in\Theta_n}\left|\log^i\left(\frac{j}{m}\right)(\frac{j}{m})^{2d}\right|\left|1-(1-\frac{1}{j})^{2d}\right| \\
& +\sup_{\theta\in\Theta_n}\left((\frac{j-1}{m})^{2d}i\left|\log^{i-1}\left(\frac{j-1}{m}\right)\frac{1}{j-1}\right|\right) \\
= & O(j^{-1}\log^i m) \text{ for all } j. \qquad \text{(A.48)}
\end{aligned}$$

Therefore

$$L_1=O_p(\frac{\log^2\lambda_m}{\sqrt{m}}\lambda_m^{2d_1}+\frac{|\log\lambda_m|\log m}{\sqrt{m}}\lambda_m^{2d_1}+\frac{\log^2 m}{\sqrt{m}}\lambda_m^{2d_1})=o(1). \qquad \text{(A.49)}$$

We then show $L_2=o_p(1)$. For $i=0,1,2$, define L_{2i} as L_{1i} is defined, but with U_j replaced by ε_j. Since $\sup_{\theta\in\Theta_n}\frac{1}{m}\sum_{j=1}^m\left|(\frac{j}{m})^{2d}\log^i(\frac{j}{m})-\frac{1}{m}\sum_{k=1}^m(\frac{k}{m})^{2d}\log^i(\frac{k}{m})\right|=O(1)$, we have

$$L_2=O_p\left(\lambda_m^{6d_0}(\log^2\lambda_m+2\log|\lambda_m|+1)\right)=o_p(1). \qquad \text{(A.50)}$$

We next show that $L_3=o_p(1)$. Following a similar procedure, we bound L_3 by $\sup_{\theta\in\Theta_n}|8\beta(d-d_0)\lambda_m^{2d}|(\log^2\lambda_m L_{31}+2|\log\lambda_m|L_{32}+L_{33})$, where

$$L_{3i+1}=\sup_{\theta\in\Theta_n}\left|\left(\frac{1}{m}\sum_{j=1}^m(\frac{j}{m})^{2d}\log^{i+1}(\frac{j}{m})\right)-\left(\frac{1}{m}\sum_{j=1}^m(\frac{j}{m})^{2d}\log^{i}(\frac{j}{m})\right)\left(\frac{1}{m}\sum_{j=1}^m\log(\frac{j}{m})\right)\right|. \qquad \text{(A.51)}$$

In view of $\frac{1}{m}\sum_{j=1}^m(\frac{j}{m})^k\log\frac{j}{m}=-\frac{1}{(k+1)^2}+o(1)$, $k\geq 0$, it is easy to show that $L_{3i+1}, i=0,1,2$ are bounded. Hence

$$L_3=O_p\left(\lambda_m^{2d_1}(\log^2\lambda_m+2|\log\lambda_m|+1)\right)=o_p(1). \qquad \text{(A.52)}$$

Continuing, we show that $L_4=o_p(1)$. Since $\sup_{\theta\in\Theta_n}\left|\beta_0\lambda_j^{2d_0}-\beta\lambda_j^{2d}\right|=O(1)$, it is easy to see that

$$L_4=O_p\left(\lambda_m^{2d_1}(\log^2\lambda_m+2|\log\lambda_m|+1)\right)=o_p(1). \qquad \text{(A.53)}$$

Therefore $\sup_{\theta\in\Theta_n}|\frac{\beta}{m}\sum_{j=1}^m e_j(\log\lambda_j^2)^2\lambda_j^{2d}|=o_p(1)$.

Following the same procedure, we can show that

$$\sup_{\theta\in\Theta_n}|\lambda_m^{-2d_0}\frac{1}{m}\sum_{j=1}^m e_j\left(\log\lambda_j^2\right)\lambda_j^{2d}| = o_p(1). \tag{A.54}$$

The details are omitted.

Part (b) We consider the individual elements of $\sup_{(d,\beta)\in\Theta_n}||D_n^{-1}[J_n(d,\beta) - J_n(d_0,\beta_0)]D_n^{-1}||$ in turn.

Since $x_{1j} = -2\log\lambda_j(1+o(1))$, the (1,1) element can be readily shown to be $o(1)$. Similarly, the (1,2) element can be written as $\sup_{(d,\beta)\in\Theta_n} 2|L_5 - L_6|(1+o(1))$ where

$$\begin{aligned} L_5 &= -\frac{1}{m}\sum_{k=1}^m\left((\frac{j}{m})^{2d} - \frac{1}{m}\sum_{k=1}^m(\frac{k}{m})^{2d}\right)\left(\log(\frac{j}{m}) - \frac{1}{m}\sum_{k=1}^m\log(\frac{k}{m})\right) \text{ and} \\ L_6 &= -\frac{1}{m}\sum_{j=1}^m\left[\left((\frac{j}{m})^{2d_0} - \frac{1}{m}\sum_{k=1}^m(\frac{k}{m})^{2d_0}\right)\left(\log(\frac{j}{m}) - \frac{1}{m}\sum_{k=1}^m\log(\frac{k}{m})\right)\right]. \end{aligned}$$

Approximating sums by integrals yields

$$\begin{aligned} L_5 &= -\frac{4d}{(2d+1)^2}(1+o(1)), \text{ and} \\ L_6 &= -\frac{4d_0}{(2d_0+1)^2}(1+o(1)). \end{aligned}$$

Therefore, the (1,2) element is $\sup_{(d,\beta)\in\Theta_n} 2|\frac{4d_0}{(2d_0+1)^2} - \frac{4d}{(2d+1)^2}|(1+o(1)) = o(1)$.

Finally, the (2,2) element is

$$\begin{aligned} &\sup_{(d,\beta)\in\Theta_n}\left|\frac{1}{m}\sum_{j=1}^m\left((\frac{j}{m})^{2d} - \frac{1}{m}\sum_{k=1}^m(\frac{k}{m})^{2d}\right)^2 - \frac{1}{m}\sum_{j=1}^m\left((\frac{j}{m})^{2d_0} - \frac{1}{m}\sum_{k=1}^m(\frac{k}{m})^{2d_0}\right)^2\right| \\ = &\sup_{(d,\beta)\in\Theta_n}\left|\frac{4d^2}{(4d+1)(2d+1)} - \frac{4d_0^2}{(4d_0+1)(2d_0+1)}\right| = o_p(1). \end{aligned} \tag{A.55}$$

Part (c) Part (c) holds by using $x_{1j} = -2\log\lambda_j(1+o(1))$ and $x_{2j} = \lambda_j^{2d_0}$ and approximating sums by integrals.

Part (d) Let $\xi_j = (\xi_{1j}, \xi_{2j})'$, where

$$\xi_{1j} = -2\log\frac{j}{m} + \frac{2}{m}\sum_{j=1}^m\log\frac{j}{m},\ \xi_{2j} = (\frac{j}{m})^{2d_0} - \frac{1}{m}\sum_{j=1}^m(\frac{j}{m})^{2d_0}. \tag{A.56}$$

Then, we can rewrite $D_n^{-1}S_n(d_0,\beta_0)$ as

$$D_n^{-1}S_n(d_0,\beta_0) = -\frac{1}{\sqrt{m}}\sum_{j=1}^m\xi_j(U_j+\varepsilon_j)(1+o(1)). \tag{A.57}$$

Note that

$$
\begin{aligned}
& \sum_{j=1}^{m} \xi_{1j}\varepsilon_j \\
= \quad & \{\sigma_u \neq \quad 0\}\lambda_m^{4d_0} \sum_{j=1}^{m} \left(-2\log\frac{j}{m} + \frac{2}{m}\sum_{j=1}^{m}\log\frac{j}{m} \right) \left(-\frac{f_w^2(0)}{2f_u^2(0)}(\frac{j}{m})^{4d_0} \right) (1+o(1)) \\
+ \quad & \{\sigma_u = \quad 0\}\lambda_m^{2} \sum_{j=1}^{m} \left(-2\log\frac{j}{m} + \frac{2}{m}\sum_{j=1}^{m}\log\frac{j}{m} \right) \left((\frac{j}{m})^2 (\frac{f_w''(0)}{2f_w(0)} + \frac{d_0}{12}) \right) (1+o(1)) \\
= \quad & \{\sigma_u \neq \quad 0\} m\lambda_m^{4d_0} \frac{f_w^2(0)}{2f_u^2(0)} \frac{8d_0}{(4d_0+1)^2} (1+o(1)) \\
- \quad & \{\sigma_u = \quad 0\} m\lambda_m^{2} \left(\frac{f_w''(0)}{f_w(0)} + \frac{d_0}{6} \right) \frac{2}{9} (1+o(1)),
\end{aligned}
\tag{A.58}
$$

and

$$
\begin{aligned}
& \sum_{j=1}^{m} \xi_{2j}\varepsilon_j \\
= \quad & \{\sigma_u \neq \quad 0\}\lambda_m^{4d_0} \sum_{j=1}^{m} \left((\frac{j}{m})^{2d_0} - \frac{1}{m}\sum_{k=1}^{m}(\frac{k}{m})^{2d_0} \right) \left(-\frac{f_u^2(0)}{2f_w^2(0)}(\frac{j}{m})^{4d_0} \right) (1+o(1)) \\
+ \quad & \{\sigma_u = \quad 0\}\lambda_m^{2} \sum_{j=1}^{m} \left((\frac{j}{m})^{2d_0} - \frac{1}{m}\sum_{k=1}^{m}(\frac{k}{m})^{2d_0} \right) \left((\frac{j}{m})^2 \left(\frac{f_w''(0)}{2f_w(0)} + \frac{d_0}{12} \right) \right) (1+o(1)) \\
= \quad & -\{\sigma_u \neq \quad 0\} m\lambda_m^{4d_0} \frac{f_w^2(0)}{2f_u^2(0)} \frac{8d_0^2}{(2d_0+1)(4d_0+1)(6d_0+1)} (1+o(1)) \\
+ \quad & \{\sigma_u = \quad 0\} m\lambda_m^{2} \left(\frac{f_w''(0)}{f_w(0)} + \frac{d_0}{6} \right) \frac{2d_0}{3\,(2d_0+3)\,(2d_0+1)} (1+o(1)).
\end{aligned}
\tag{A.59}
$$

Therefore

$$
D_n^{-1} S_n(d_0, \beta_0) + b_n = \frac{1}{\sqrt{m}} \sum_{j=1}^{m} \xi_j U_j + o(1). \tag{A.60}
$$

We now prove that for any vector $v = (v_1, v_2)'$, $\frac{1}{\sqrt{m}} \sum_{j=1}^{m} v'\xi_j U_j \Rightarrow N(0, \frac{\pi^2}{6} v'\Omega v)$. Write

$$
\frac{1}{\sqrt{m}} \sum_{j=1}^{m} v'\xi_j U_j = T_1 + T_2 + T_3, \tag{A.61}
$$

where

$$T_1 = \frac{1}{\sqrt{m}} \sum_{j=1}^{\log^8 m} a_j U_j, T_2 = \frac{1}{\sqrt{m}} \sum_{j=\log^8 m+1}^{m^{0.5+\delta}} a_j U_j$$
$$T_3 = \frac{1}{\sqrt{m}} \sum_{j=m^{0.5+\delta}}^{m} a_j U_j, a_j = v'\xi_j, \quad \text{(A.62)}$$

for some $0 < \delta < 0.5$.

Since $\max_{1\leq j\leq m} |\xi_{1j}| = O(\log m)$ and $\max_{1\leq j\leq m} |\xi_{2j}| = O(\log m)$, we have $\max_{1\leq j\leq m} |a_j| = O(\log m)$. Therefore the proofs in HDB that $T_1 = o_p(1)$ and $T_2 = o_p(1)$ are also valid in the present case. We now show that $T_3 \to N(0, \frac{\pi^2}{6} v'\Omega v)$ by verifying that the sequence $\{a_j\}$ satisfies (24) with $\rho = v'\Omega v$. The first condition of (24) holds as $\max_{1\leq j\leq m} |a_j| = O(\log m) = o(m)$. The second condition holds because

$$\sum_{j=m^{0.5+\delta}+1}^{m} a_j^2 = \sum_{j=1}^{m} a_j^2 - \sum_{j=1}^{m^{0.5+\delta}} a_j^2 = \sum_{j=1}^{m} a_j^2 + o(m)$$
$$= mv'(\frac{1}{m}\sum_{j=1}^{m} \xi_j'\xi_j)v + o(m) \sim mv'\Omega v. \quad \text{(A.63)}$$

The last equality follows because we can show that $\lim_{m\to\infty} \frac{1}{m}\sum_{j=1}^{m} \xi_j'\xi_j = \Omega$ by approximating the sums by integrals. The third condition holds because

$$\sum_{j=m^{0.5+\delta}+1}^{m} |a_j|^p$$
$$\leq 2^p|v_1| \sum_{j=m^{0.5+\delta}+1}^{m} |\xi_{1j}|^p + 2^p|v_2| \sum_{j=m^{0.5+\delta}+1}^{m} |\xi_{2j}|^p$$
$$= O(m) + 2^p|v_1| \sum_{j=m^{0.5+\delta}+1}^{m} |\xi_{2j}|^p$$
$$= O(\sum_{j=m^{0.5+\delta}+1}^{m} |(\frac{2\pi j}{m})^{2d_0}|^p) + O\left\{\sum_{j=m^{0.5+\delta}+1}^{m} \left[\frac{1}{m}\sum_{j=1}^{m}(\frac{2\pi j}{m})^{2d_0}\right]^p\right\} + O(m)$$
$$= O(m) + O(m) + O(m) = O(m). \quad \text{(A.64)}$$

Here we have employed $\sum_{j=m^{0.5+\delta}+1}^{m} |\xi_{1j}|^p = O(m)$. See $(A18)$ in HDB (1998).

The above results combine to establish part (d). □

Proof of theorem 3

Scaling the first order conditions, we have

$$-D_n^{-1}S_n(d_0,\beta_0) = D_n^{-1}H_n(d_0,\beta_0)D_n^{-1}D_n(\widehat{d}-d_0,\widehat{\beta}-\beta_0)'$$
$$+D_n^{-1}[H_n^* - H_n(d_0,\beta_0)]D_n^{-1}D_n(\widehat{d}-d_0,\widehat{\beta}-\beta_0)'. \quad \text{(A.65)}$$

Thus

$$
\begin{aligned}
& D_n(\widehat{d}-d_0, \widehat{\beta}-\beta_0)' & \text{(A.66)} \\
= & -\left\{D_n^{-1} H_n(d_0,\beta_0) D_n^{-1} + D_n^{-1}[H_n^* - H_n(d_0,\beta_0)] D_n^{-1}\right\}^{-1} D_n^{-1} S_n(d_0,\beta_0).
\end{aligned}
$$

But since $\widehat{d}-d_0 = O_p((\frac{m}{n})^{2d_0})$, we know that $(\widehat{d},\widehat{\beta})$ and (d^*,β^*) belong to Θ_n with probability approaching one. Therefore,

$$
\begin{aligned}
& ||D_n^{-1}[H_n^* - H_n(d_0,\beta_0)] D_n^{-1}|| \\
\leq & \sup_{(d,\beta)\in\Theta_n} \left(||D_n^{-1}[H_n(d,\beta) - J_n(d,\beta)] D_n^{-1}|| + ||D_n^{-1}[H_n(d_0,\beta_0) - J_n(d_0,\beta_0)] D_n^{-1}||\right) \\
& + \sup_{(d,\beta)\in\Theta_n} ||D_n^{-1}[J_n(d,\beta) - J_n(d_0,\beta_0)] D_n^{-1}|| \\
= & o_p(1), \qquad \text{(A.67)}
\end{aligned}
$$

by Lemma 5. Furthermore,

$$
\begin{aligned}
D_n^{-1} H_n(d_0,\beta_0) D_n^{-1} &= D_n^{-1}[H_n(d_0,\beta_0) - J_n(d_0,\beta_0)] D_n^{-1}|| + D_n^{-1} J_n(d_0,\beta_0) D_n^{-1} \\
&= \Omega + o(1). \qquad \text{(A.68)}
\end{aligned}
$$

Consequently,

$$
\begin{aligned}
D_n(\widehat{d}-d_0, \widehat{\beta}-\beta_0)' - \Omega^{-1} b_n &= -\Omega^{-1}\left(D_n^{-1} S_n(d_0,\beta_0) + b_n\right) + o_p(1) \\
&\Rightarrow -\Omega^{-1} N(0, \frac{\pi^2}{6}\Omega) =_d N(0, \frac{\pi^2}{6}\Omega^{-1}). \qquad \text{(A.69)}
\end{aligned}
$$

□

Figure 1. Performance of the new estimator and the GPH estimator with $d_0 = 0.45$, $\sigma_u^2 = 0$ for sample size 512, computed using 2000 repetitions.

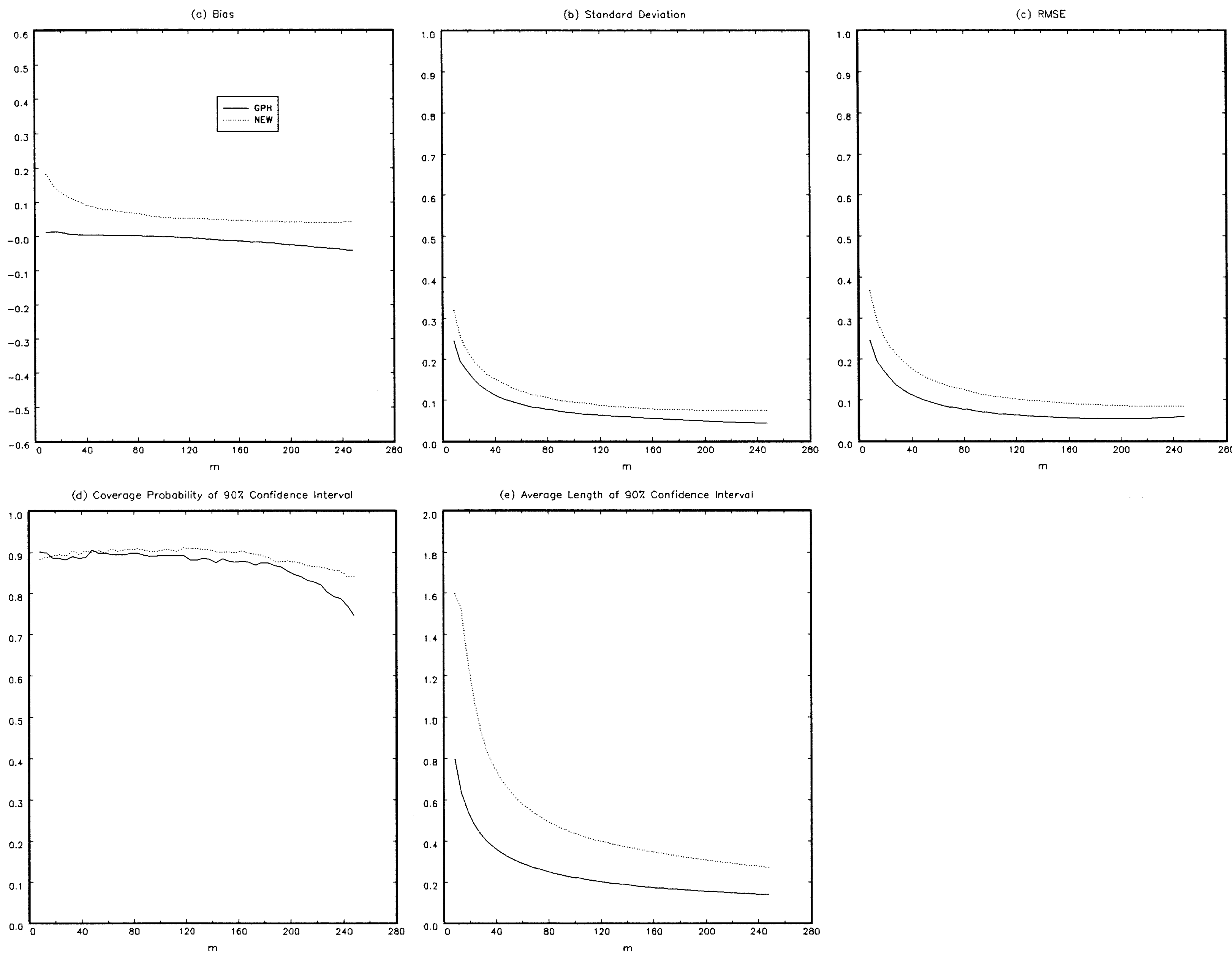


Figure 2. Performance of the new estimator and the GPH estimator with $d_0 = 0.45$, $\sigma_u^2 = 8$ for sample size 512, computed using 2000 repetitions.

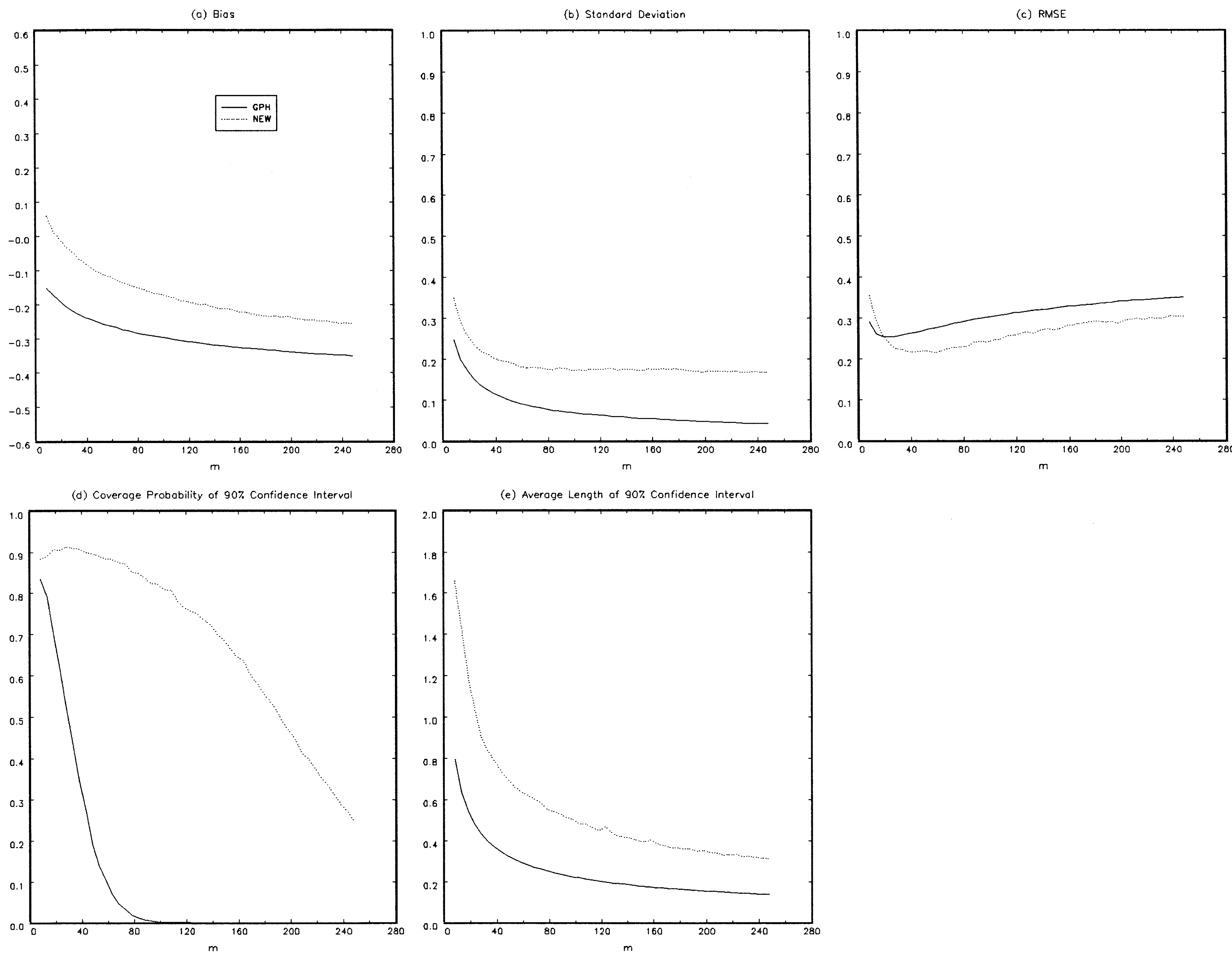


Figure 3. Performance of the New Estimator and the GPH estimator with $d_0 = 0.85$, $\sigma_u^2 = 0$ for sample size 512, computed using 2000 repetitions.

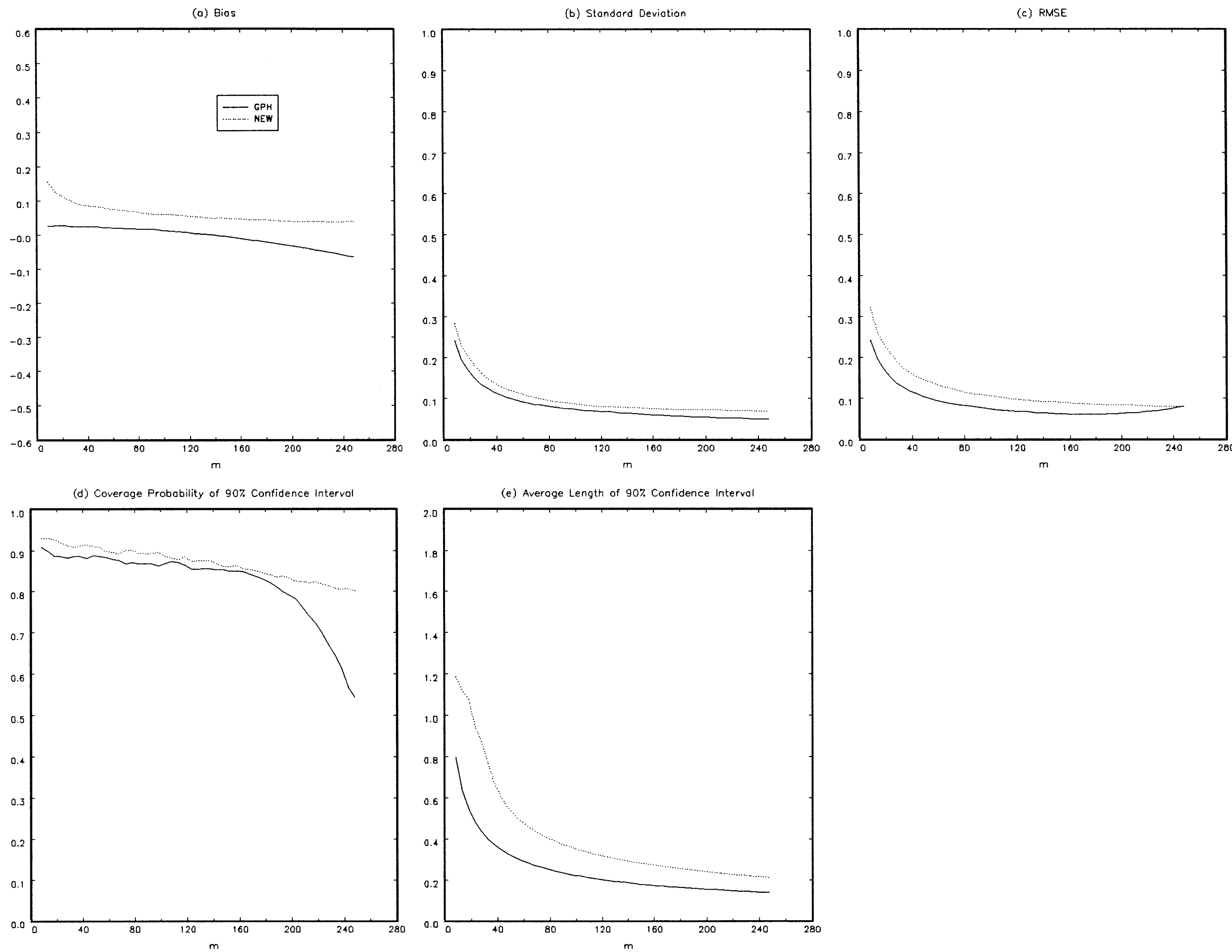


Figure 4. Performance of the new estimator and the GPH estimator with $d_0 = 0.85$, $\sigma_u^2 = 8$ for sample size 512, computed using 2000 repetitions.

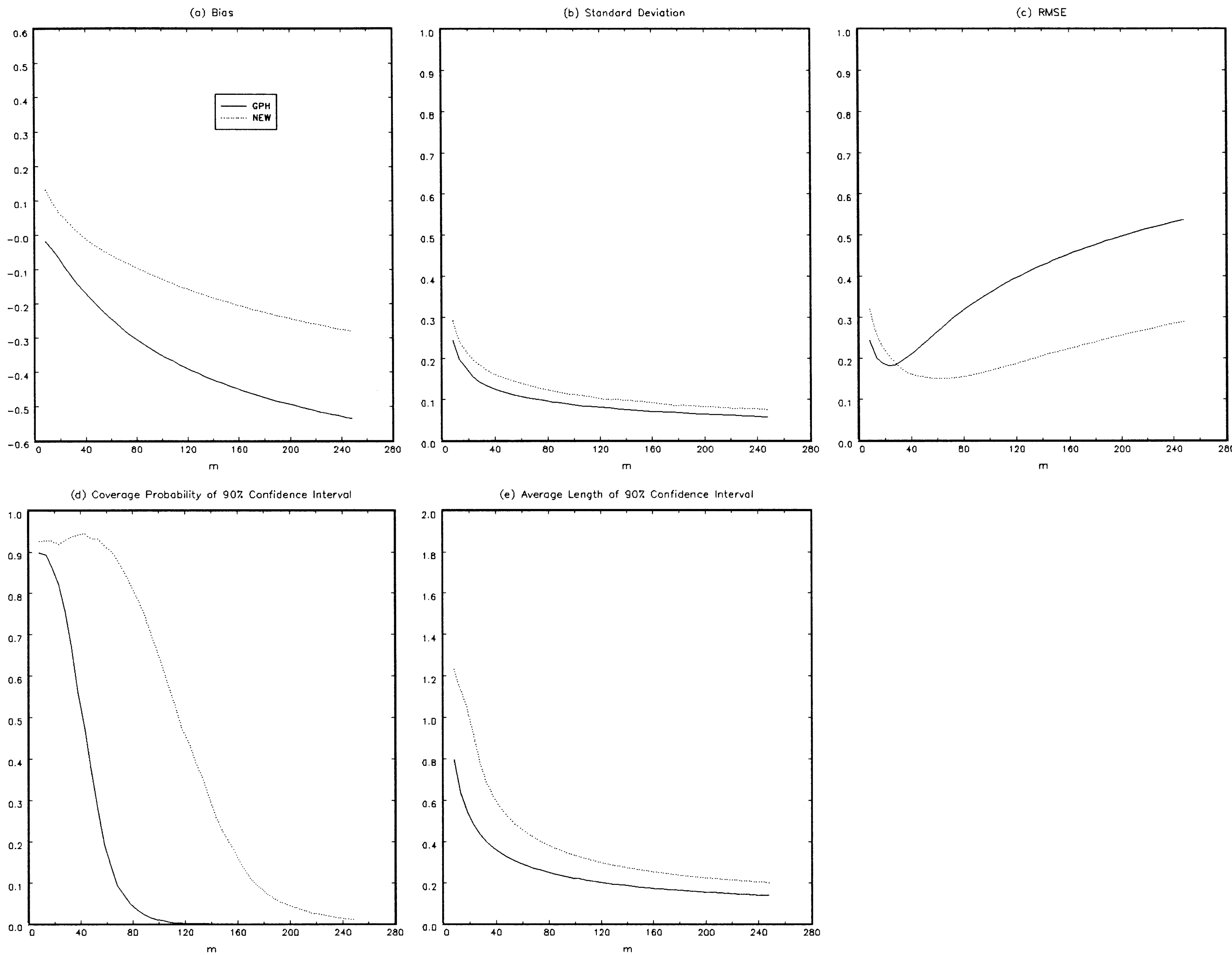
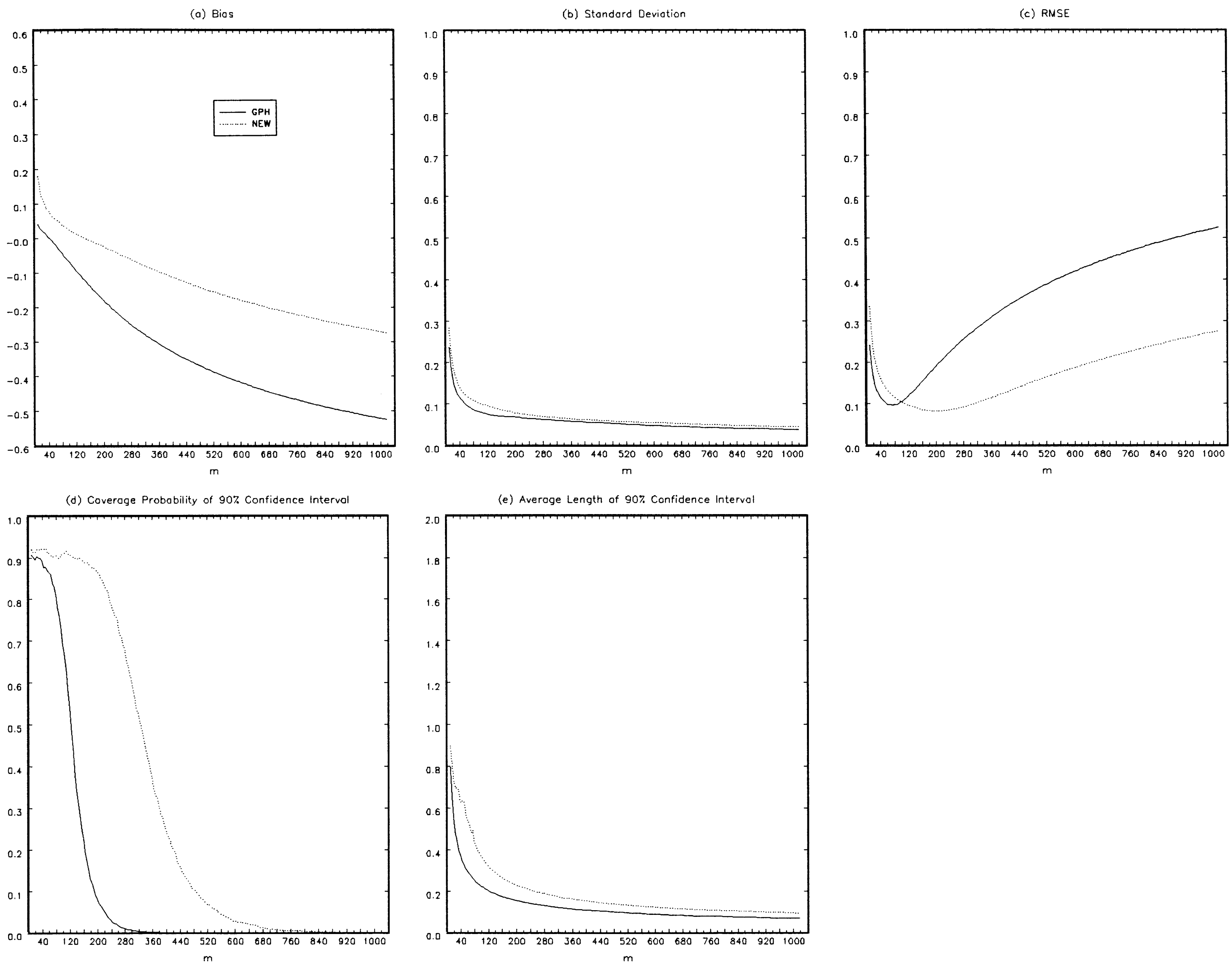


Figure 5. Performance of the new estimator and the GPH estimator with $d_0 = 0.85$, $\sigma_u^2 = 8$ for sample size 2048, computed using 2000 repetitions.



References

Andrews, D. W. K. and P. Guggenberger, 1999. A bias-reduced log-periodogram regression estimator for the long-memory parameter. Cowles Foundation Discussion Paper No. 1263, Yale University.

Andrews, D. W. K. and Y. Sun, 2000. Local polynomial Whittle estimation of long range dependence. Cowles Foundation Discussion Paper No. 1293, Yale University.

Breidt, F. J. , N. Crato and P. De Lima, 1998. The detection and estimation of long memory in stochastic volatility. Journal of Econometrics 83, 32-348.

Deo R. S. and C. M. Hurvich, 1999. On the log periodogram regression estimator of the long memory parameter in the long memory stochastic volatility models. New York University.

Geweke, J., and S. Porter-Hudak, 1983. The estimation and application of long memory time series models. Journal of Time Series Analysis 4, 221-38.

Granger C. W. J. and F. Marmol, 1997. The correlogram of a long memory process plus a simple noise. Department of Economics, University of California, San Diego.

Henry, M. and P. M. Robinson, 1996. Bandwidth choice in Gaussian semiparametric estimation of long range dependence. in: P. M. Robinson and M. Rosenblatt, eds., Athens Conference on Applied Probability and Time Series in memory of E. J. Hannan: Volume II.

Henry, M., 1999. Robust automatic bandwidth for long memory. Department of Economics, Columbia University.

Hurvich, C. M. and R. S. Deo, 1999. Plug-in selection of the number of frequencies in regression estimates of the memory parameter of a long memory time series. Journal of Time Series Analysis 20, 331-341.

Hurvich, C. M. and J. Brodsky, 1997. Broadband semiparametric estimation of the memory parameter of a long memory time series using fractional exponential models. Department of Statistics and Operations Research, New York University.

Hurvich, C. M., R. S. Deo and J. Brodsky, 1998. The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long memory time series. Journal of Time Series Analysis 19, 19-46.

Kim C. and P. C. B. Phillips, 2000. Modified log periodogram regression. Department of Economics, Yale University.

Künsch, H., 1986. Discrimination between monotonic trends and long-range dependence. Journal of Applied Probability 23, 1025-30.

Linton, O. and P. Gozalo, 2000. Local nonlinear least squares: using parametric information in nonparametric regression. Journal of Econometrics 99, 63-106.

Moulines, E. and P. Soulier, 1999. Broadband log-periodogram regression of time series with long-range dependence. Annals of Statistics 27, 1415-1439.

Phillips, P. C. B., 1998. Unit root log-periodogram regression. Department of Economics, Yale University.

Phillips, P. C. B., 1999. Discrete Fourier transforms of fractional processes. Cowles Foundation, Yale University.

Robinson, P. M., 1995a. Log-periodogram regression of time series with long range dependence. Annals of Statistics 23, 1048-1072.

Robinson, P. M., 1995b. Gaussian semiparametric estimation of long range dependence. Annals of Statistics 23, 1630–1661.

Shimotsu, K. and P. C. B. Phillips, 2001. Exact local Whittle estimation of fractional integration. Cowles Foundation, Yale University.

Sun, Y., and P. C. B. Phillips, 2000. Perturbed fractional process, fractional cointegration and the Fisher hypothesis. Department of Economics, Yale University.